



**University of
Zurich**^{UZH}

**Zurich Open Repository and
Archive**

University of Zurich
University Library
Strickhofstrasse 39
CH-8057 Zurich
www.zora.uzh.ch

Year: 2017

Autoregressive Lag-Order Selection Using Conditional Saddlepoint Approximations

Butler, Ronald W ; Paoella, Marc S

Abstract: A new method for determining the lag order of the autoregressive polynomial in regression models with autocorrelated normal disturbances is proposed. It is based on a sequential testing procedure using conditional saddlepoint approximations and permits the desire for parsimony to be explicitly incorporated, unlike penalty-based model selection methods. Extensive simulation results indicate that the new method is usually competitive with, and often better than, common model selection methods.

DOI: <https://doi.org/10.3390/econometrics5030043>

Posted at the Zurich Open Repository and Archive, University of Zurich

ZORA URL: <https://doi.org/10.5167/uzh-142227>

Journal Article

Published Version



The following work is licensed under a Creative Commons: Attribution 4.0 International (CC BY 4.0) License.

Originally published at:

Butler, Ronald W; Paoella, Marc S (2017). Autoregressive Lag-Order Selection Using Conditional Saddlepoint Approximations. *Econometrics*, 5(3):43.

DOI: <https://doi.org/10.3390/econometrics5030043>

Article

Autoregressive Lag—Order Selection Using Conditional Saddlepoint Approximations

Ronald W. Butler¹ and Marc S. Paoletta^{2,*}¹ Department of Statistical Science, Southern Methodist University, Dallas, TX 75275-0332, USA; rbutler@smu.edu² Department of Banking and Finance, University of Zurich, Zurich 8032, Switzerland

* Correspondence: marc.paoletta@bf.uzh.ch; Tel.: 41-44-634-4584

Academic Editors: Federico Bandi, Alex Maynard, Hyungsik Roger Moon and Benoit Perron

Received: 8 May 2017; Accepted: 6 September 2017; Published: 19 September 2017

Abstract: A new method for determining the lag order of the autoregressive polynomial in regression models with autocorrelated normal disturbances is proposed. It is based on a sequential testing procedure using conditional saddlepoint approximations and permits the desire for parsimony to be explicitly incorporated, unlike penalty-based model selection methods. Extensive simulation results indicate that the new method is usually competitive with, and often better than, common model selection methods.

Keywords: ARMA; saddlepoint approximation; simplicity

JEL Classification: C1

1. Introduction

We consider the performance of a new method for selecting the appropriate lag order p of an autoregressive (AR) model for the residuals of a linear regression. Our focus is on the small-sample performance as compared to competing methods, and, as such, we concentrate on the underlying small-sample distribution theory employed, instead of considerations of consistency and performance in an asymptotic framework.

The problem of ARMA model specification has a long history, and we omit a literature review, though note the book by Choi (1992), dedicated to the topic. Less well documented is the use of small-sample distributional approximations using saddlepoint techniques. Saddlepoint methodology started with the seminal contributions from Daniels (1954,1987) and continued with those from Barndorff-Nielsen and Cox (1979), Skovgaard (1987), Reid (1988), Daniels and Young (1991) and Kolassa (1996). It has been showcased in the book-length treatments of Field and Ronchetti (1990), Jensen (1995), Kolassa (2006), and Butler (2007), the latter showing the enormous variety of problems in statistical inference that are amenable to its use.

The first uses of saddlepoint approximations for inference concerning serial correlation are Daniels (1956) and McGregor (1960). This was followed by Phillips (1978), who also used such methods for simultaneous systems in Holly and Phillips (1979). Saddlepoint approximations have also been used for computing confidence intervals and unit root inference in first-order autoregressive models; see e.g., Broda et al. (2007), which builds on the methodology in Andrews (1993); and Broda et al. (2009). Further work on construction of confidence intervals in the near unit-root case can be found in Elliott and Stock (2001), Andrews and Guggenberger (2014) and Phillips (2014).

Related work can be found in [Ploberger and Phillips \(2003\)](#), [Leeb and Pötscher \(2005\)](#) and [Phillips \(2008\)](#). Peter Phillips continues to work on the selection of the autoregressive order; see e.g., [Han et al. \(2017\)](#) for its study in the panel data case.

In our setting herein, we restrict attention to the stationary setting and do not explicitly consider the unit root case. Time series $\mathbf{Y} = (Y_1, \dots, Y_T)'$ is assumed to have distribution $\mathbf{Y} \sim N(\mathbf{X}\boldsymbol{\beta}, \sigma^2\boldsymbol{\Psi}^{-1})$, where \mathbf{X} is a full rank, $T \times k$ matrix of exogenous variables and $\boldsymbol{\Psi}^{-1}$ is the covariance matrix corresponding to a stationary AR(p) model with parameter vector $\boldsymbol{\phi} = (\phi_1, \dots, \phi_p)$. Values $\boldsymbol{\beta}$, σ^2 , $\boldsymbol{\phi}$ and p are fixed but unknown. Below (in Section 6), we extend to an ARMA framework, such that, in addition, the MA parameters $\boldsymbol{\theta} = (\theta_1, \dots, \theta_q)$ are unknown, but q is assumed known.

The two most common approaches used to determine the appropriate AR lag order are (i) assessing the lag at which the sample partial autocorrelation “cuts off”; and (ii) use of popular penalty-based model selection criteria such as AIC and SBC/BIC (see, e.g., [Konishi and Kitagawa 2008](#); [Brockwell and Davis 1991](#), *sect. 9.3*; [McQuarrie and Tsai 1998](#); [Burnham and Anderson 2003](#)). As emphasized in [Chatfield \(2001, pp. 31, 91\)](#), interpreting a sample correlogram is one of the hardest tasks in time series analysis, and their use for model determination is, at best, difficult, and often involves considerable experience and subjective judgment. The fact that the distribution of the sample autocorrelation function (SACF) and the sample partial ACF (SPACF) of the regression residuals are—especially for smaller sample sizes—highly dependent on the \mathbf{X} matrix makes matters significantly more complicated. In comparison, the application of penalty-based model selection criteria is virtually automatic, requiring only the choice of which criteria to use. The criticism that their use involves model estimation and, thus, far more calculation is, with modern computing power, no longer relevant.

A different, albeit seldom used, identification methodology involves sequential testing procedures. One approach sequentially tests

$$H_m : \phi_m = 0, \quad H_{m-1} : \phi_m = \phi_{m-1} = 0, \dots, \quad H_1 : \phi_m = \dots = \phi_1 = 0.$$

Testing stops when the first hypothesis is rejected (and all remaining are then also rejected). Tests of the k th null hypothesis can be based on a scaled sum of squared values of the SPACF or numerous variations thereof, all of which are asymptotically χ^2 distributed; see [Choi \(1992, chp. 6\)](#) and the references therein for a detailed discussion.

The use of sequential testing procedures in this context is not new. For example, [Jenkins and Alevi \(1981\)](#) and [Tiao and Box \(1981\)](#) propose methods based on the asymptotic distribution of the SPACF under the null of white noise. More generally, [Pötscher \(1983\)](#) considers determination of optimal values of p and q by a sequence of Lagrange multiplier tests. In particular, for a given choice of maximal orders, P and Q , and a chain of (p, q) -values $(p_0, q_0) = (0, 0)$, $(p_1, q_1), \dots, (p_K, q_K) = (P, Q)$, such that either $p_{i+1} = p_i$ and $q_{i+1} = q_i + 1$ or $p_{i+1} = p_i + 1$ and $q_{i+1} = q_i$, $i = 0, 1, \dots, K = P + Q$, a sequence of Lagrange-multiplier tests are performed for each possible chain. The optimal orders are obtained when the test does not reject for the first time. As noted by [Pötscher \(1983, p. 876\)](#), “strong consistency of the estimators is achieved if the significance levels of all the tests involved tend to zero with increasing [sample] size...”. This forward search procedure is superficially similar to the method proposed herein, and also requires specification of a sequence of significance levels. Our method differs in two important regards. First, near-exact small-sample distribution theory is employed by use of conditional

saddlepoint approximations; and second, we explicitly allow for, and account for, a mean term in the form of a regression $X\beta$.

Besides the inherent problems involved in sequential testing procedures, such as controlling overall type I errors and the possible tendency to over-fit (as in backward regression procedures), the reliance on asymptotic distributions can be detrimental when working with small samples and unknown mean term $X\beta$. This latter problem could be overcome by using exact small-sample distribution theory and a sequence of point optimal tests in conjunction with some prior knowledge of the AR coefficients; see, e.g., [King and Srikanthakumar \(2015\)](#) and the references therein.

Compared to sequential hypothesis testing procedures, penalty-based model selection criteria have the arguable advantage that there is no model preference via a null hypothesis, and that the order in which calculations are performed is irrelevant (see, for example, [Granger et al. 1995](#)). On the other hand, as is forcefully and elegantly argued in [Zellner \(2001\)](#) in a general econometric modeling context, it is worthwhile to have an ordering of possible models in terms of complexity, with higher probabilities assigned to simpler models. Moreover, [Zellner \(2001, sct. 3\)](#) illustrates the concept with the choice of ARMA models, discouraging the use of MA components in favor of pure AR processes, even if it entails more parameters, because “counting parameters” is not necessarily a measure of complexity (see also [Keuzenkamp and McAleer 1997](#), p. 554). This agrees precisely with the general findings of [Makridakis and Hibon \(2000, p. 458\)](#), who state “Statistically sophisticated or complex models do not necessarily produce more accurate forecasts than simpler ones”. With such a modeling approach, the aforementioned disadvantages of sequential hypothesis testing procedures become precisely its advantages. In particular, one is able to specify error rates on individual hypotheses corresponding to models of differing complexity.

In this paper, we present a sequential hypothesis testing procedure for computing the lag length that, in comparison to the somewhat ad hoc sequential methods mentioned above, operationalizes the uniformly most powerful unbiased (UMPU) test of [Anderson \(1971, pp. 34–46, 260–76\)](#). It makes use of a conditional saddlepoint approximation to the—otherwise completely intractable—distribution of the m th sample autocorrelation given those of order $1, 2, \dots, m - 1$. While exact calculation of the required distribution is not possible (not even straightforward via simulation because of the required conditioning), Section 4 provides evidence that the saddlepoint approximation is, for practical purposes, virtually exact in this context.

The remainder of the paper is outlined as follows. Sections 2 and 3 briefly outline the required distribution theory of the sample autocorrelation function and the UMPU test, respectively. Section 4 illustrates the performance of the proposed method in the null case of no autocorrelation, while Section 5 compares the performance of the new and existing order selection methods for several autoregressive structures. Section 6 proposes an extension of the method to handle AR lag selection in the presence of ARMA disturbances. Section 7 provides a performance comparison when the Gaussianity assumption is violated. Section 8 provides concluding remarks.

2. The Distribution of the Autocorrelation Function

Define the $T \times T$ matrix \mathbf{A}_l such that its (i, j) -th element is given by $a_{ij} = \mathbb{I}\{|i - j| = l\} / 2$ and $\mathbb{I}\{\cdot\}$ denotes the indicator function. Then, for covariance stationary, mean-zero vector $\boldsymbol{\epsilon} = (\epsilon_1, \dots, \epsilon_T)'$, the ratio of quadratic forms

$$R_l = \frac{\boldsymbol{\epsilon}' \mathbf{A}_l \boldsymbol{\epsilon}}{\boldsymbol{\epsilon}' \boldsymbol{\epsilon}} = \frac{\hat{\gamma}(l)}{\hat{\gamma}(0)}, \quad \hat{\gamma}(l) = T^{-1} \sum_{t=1}^{T-l} \epsilon_t \epsilon_{t+l},$$

is the usual estimator of the l th lag autocorrelation coefficient. The sample autocorrelation function with m components, hereafter SACF, is then given by vector $R_{(m)} = (R_1, \dots, R_m)$ with joint density denoted by $f_{R_{(m)}}$.

Recall that a function $\gamma : \mathbb{Z} \rightarrow \mathbb{R}$ is the autocovariance function of a weakly stationary time series if and only if γ is even, i.e., $\gamma(h) = \gamma(-h)$ for all $h \in \mathbb{Z}$, and is positive semi-definite. See, e.g., [Brockwell and Davis \(1991, p. 27\)](#) for proof. Next recall that a symmetric matrix is positive definite if and only if all the leading principal minors are positive; see, e.g., [Abadir and Magnus \(2005, p. 223\)](#). As such, the support of $R_{(m)}$ is given by

$$\mathfrak{S}_m = \{\mathbf{r} = (r_1, r_2, \dots, r_m) : |\mathbf{R}_i| > 0, \quad i = 1, \dots, m\}, \quad (1)$$

where \mathbf{R}_i is the band matrix given by

$$\mathbf{R}_i = \begin{pmatrix} 1 & r_1 & \cdots & r_i \\ r_1 & 1 & \cdots & r_{i-1} \\ \vdots & & \ddots & \vdots \\ r_i & \cdots & r_1 & 1 \end{pmatrix}. \quad (2)$$

Assume for the moment that there are no regression effects and let $\boldsymbol{\epsilon} \sim N_T(\mathbf{0}, \boldsymbol{\Omega}^{-1})$ with $\boldsymbol{\Omega}^{-1} > 0$, i.e., $\boldsymbol{\Omega}$ is positive definite. While no tractable expression for $f_{R_{(m)}}$ appears to exist, a saddlepoint approximation to the density of $R_{(m)}$ at $\mathbf{r} = (r_1, \dots, r_m)'$ is shown in [Butler and Paolella \(1998\)](#) to be given by

$$\hat{f}_{R_{(m)}}(\mathbf{r}) = (2\pi)^{-\frac{m}{2}} |\boldsymbol{\Omega}|^{\frac{1}{2}} |\hat{\mathbf{H}}_{\boldsymbol{\Omega}}|^{-\frac{1}{2}} |\hat{\mathbf{P}}_{\boldsymbol{\Omega}}|^{-\frac{1}{2}} (\text{tr } \hat{\mathbf{P}}_{\boldsymbol{\Omega}}^{-1})^m, \quad (3)$$

where

$$\hat{\mathbf{P}}_{\boldsymbol{\Omega}} = \hat{\mathbf{P}}_{\boldsymbol{\Omega}}(\hat{\mathbf{s}}) = \boldsymbol{\Omega} + 2\mathbf{r}'\hat{\mathbf{s}}\mathbf{I}_T - 2\sum_{i=1}^m \hat{s}_i \mathbf{A}_i,$$

and $\hat{\mathbf{H}}_{\boldsymbol{\Omega}} = \hat{\mathbf{H}}_{\boldsymbol{\Omega}}(\hat{\mathbf{s}})$ with (i, j) -th element given by

$$\hat{h}_{ij} = -\frac{1}{2} \frac{\partial^2}{\partial \hat{s}_i \partial \hat{s}_j} \log(|\hat{\mathbf{P}}_{\boldsymbol{\Omega}}(\hat{\mathbf{s}})|) = 2 \text{tr} \left[\hat{\mathbf{P}}_{\boldsymbol{\Omega}}^{-1} (\mathbf{A}_i - r_i \mathbf{I}_T) \hat{\mathbf{P}}_{\boldsymbol{\Omega}}^{-1} (\mathbf{A}_j - r_j \mathbf{I}_T) \right], \quad (4)$$

$i, j = 1, \dots, m$. Saddlepoint vector $\hat{\mathbf{s}} = (\hat{s}_1, \dots, \hat{s}_m)'$ solves

$$0 = -\frac{1}{2} \frac{\partial}{\partial \hat{s}_i} \log(|\hat{\mathbf{P}}_{\boldsymbol{\Omega}}(\hat{\mathbf{s}})|) = \text{tr} \left[\hat{\mathbf{P}}_{\boldsymbol{\Omega}}^{-1} (\mathbf{A}_i - r_i \mathbf{I}_T) \right], \quad i = 1, \dots, m, \quad (5)$$

and, in general, needs to be numerically obtained. In the null setting for which $\Omega = \mathbf{I}_T$, $\text{tr}\hat{\mathbf{P}}_T^{-1} = T$, so that the last factor in (3) is just T^m . Special cases for which explicit solutions to (5) exist are given in Butler and Paoletta (1998).

With respect to the calculation of Ω corresponding to a stationary, invertible ARMA(p, q) process, the explicit method in McLeod (1975) could be used, though more convenient in matrix-based computing platforms are the matrix expressions given in Mittnik (1988) and Van der Leeuw (1994). Code for the latter, as well as for computing the CACF test, are available upon request.

The extension of (3) for use with regression residuals is not immediately possible because the covariance matrix of $\hat{\epsilon}$ is not full rank and a canonical reduction of the residual vector is required. To this end, let

$$\mathbf{y} \sim N(\mathbf{X}\beta, \Psi^{-1}), \quad (6)$$

where \mathbf{X} is a full rank $T \times k$ matrix of exogenous variables. Denote the OLS residual vector as $\hat{\epsilon} = \mathbf{M}\mathbf{y} = \mathbf{M}\epsilon$, where $\mathbf{M} = \mathbf{I}_T - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$. As \mathbf{M} is an orthogonal projection matrix, it can be expressed as $\mathbf{M} = \mathbf{G}'\mathbf{G}$, where \mathbf{G} is $(T - k) \times T$ and such that $\mathbf{G}\mathbf{G}' = \mathbf{I}_{T-k}$ and $\mathbf{G}\mathbf{X} = \mathbf{0}$. Then

$$R_l = \frac{\hat{\epsilon}'\mathbf{A}_l\hat{\epsilon}}{\hat{\epsilon}'\hat{\epsilon}} = \frac{\mathbf{y}'\mathbf{M}\mathbf{A}_l\mathbf{M}\mathbf{y}}{\mathbf{y}'\mathbf{M}\mathbf{y}} = \frac{\mathbf{y}'\mathbf{G}'\mathbf{G}\mathbf{A}_l\mathbf{G}'\mathbf{G}\mathbf{y}}{\mathbf{y}'\mathbf{G}'\mathbf{G}\mathbf{y}} = \frac{\mathbf{w}'\tilde{\mathbf{A}}_l\mathbf{w}}{\mathbf{w}'\mathbf{w}}, \quad (7)$$

where $\mathbf{w} = \mathbf{G}\mathbf{y}$ and $\tilde{\mathbf{A}}_l = \mathbf{G}\mathbf{A}_l\mathbf{G}'$ is a $(T - k) \times (T - k)$ symmetric matrix. For example, \mathbf{G} could consist of an orthogonal basis for the $T - k$ eigenvectors of \mathbf{M} corresponding to the unit eigenvalues. By setting $\Omega^{-1} = \mathbf{G}\Psi^{-1}\mathbf{G}'$, approximation (3) becomes valid using $\mathbf{w} \sim N(\mathbf{0}, \Omega^{-1})$ and $\mathbf{G}\mathbf{A}_l\mathbf{G}'$ in place of ϵ and \mathbf{A}_l , respectively. Note that, in the null case with $\mathbf{y} \sim N(\mathbf{X}\beta, \sigma^2\mathbf{I}_T)$, $\Omega^{-1} = \sigma^2\mathbf{I}_{T-k}$.

3. Conditional Distributions and UMPU Tests

Anderson (1971, sct. 6.3.2) has shown for the regression model with circular AR(m) errors (so $\epsilon_1 \equiv \epsilon_T$) and the columns of \mathbf{X} restricted to Fourier regressors, i.e.,

$$y_t = \beta_1 + \sum_{s=1}^{(k-1)/2} \left\{ \beta_{2s} \cos\left(\frac{2\pi st}{T}\right) + \beta_{2s+1} \sin\left(\frac{2\pi st}{T}\right) \right\} + \epsilon_t, \quad (8)$$

that the uniformly most powerful unbiased (UMPU) test of AR($m - 1$) versus AR(m) disturbances rejects for values of r_m falling sufficiently far out in either tail of the conditional density

$$f_{R_m|R_{(m-1)}}(r_m | \mathbf{r}_{(m-1)}), \quad (9)$$

where $\mathbf{r}_{(m-1)} = (r_1, \dots, r_{m-1})'$ denotes the observed value of the vector of random variables $R_{(m-1)}$. A p -value can be computed as $\min\{\tau_m, 1 - \tau_m\}$, where

$$\tau_1 = \Pr(R_1 < r_1) \quad \text{and} \quad \tau_m = \Pr(R_m < r_m | R_{(m-1)} = \mathbf{r}_{(m-1)}), \quad m > 1. \quad (10)$$

Like in the well-studied $m = 1$ case (cf. (Durbin and Watson 1950, 1971) and the references therein), the optimality of the test breaks down in either the non-circular model and/or with arbitrary exogenous \mathbf{X} , but does provide strong motivation for an approximately UMPU test in the general setting considered here. This is particularly so for economic time series, as they typically exhibit

seasonal (i.e., cyclical) behavior that can mimic the Fourier regressors in (8) (Dubbelman et al. 1978; King 1985, p. 32).

Following the methodology outlined in Barndorff-Nielsen and Cox (1979), Butler and Paoletta (1998) derive a conditional double saddlepoint density to (9) computed as the ratio of two single approximations

$$\begin{aligned}\hat{f}_{R_m|R_{(m-1)}}(r_m | \mathbf{r}_{(m-1)}; \Omega) &= \frac{\hat{f}_{R_{(m)}}(\mathbf{r}_{(m-1)}, r_m; \Omega)}{\hat{f}_{R_{(m-1)}}(\mathbf{r}_{(m-1)}; \Omega)} \\ &= \sqrt{\frac{|\hat{\mathbf{H}}_{m-1}| |\hat{\mathbf{P}}_{m-1}|}{2\pi |\hat{\mathbf{H}}_m| |\hat{\mathbf{P}}_m|}} \left(\text{tr} \hat{\mathbf{P}}_m^{-1} \right)^m \left(\text{tr} \hat{\mathbf{P}}_{m-1}^{-1} \right)^{1-m},\end{aligned}\quad (11)$$

where $\hat{\mathbf{H}}_{m-1}$ and $\hat{\mathbf{P}}_{m-1}$ are the $\hat{\mathbf{H}}_\Omega$ and $\hat{\mathbf{P}}_\Omega$ values, respectively, associated with the $(m-1)$ -dimensional saddlepoint $\hat{\mathbf{s}}_{m-1}$ of the denominator determined by $\mathbf{r}_{(m-1)}$; likewise for $\hat{\mathbf{H}}_m$ and $\hat{\mathbf{P}}_m$, and explicit dependence on Ω has been suppressed.

Thus, τ_m in (10) for $m > 1$ can be computed as

$$\tau_m = \frac{\int_{\mathfrak{S}'_m} \hat{f}_{R_m|R_{(m-1)}}(r) \, \mathrm{d}r}{\int_{\mathfrak{S}_m} \hat{f}_{R_m|R_{(m-1)}}(r) \, \mathrm{d}r} = \frac{\int_{\mathfrak{S}'_m} \hat{f}_{R_{(m)}}(\mathbf{r}_{(m-1)}, r_m; \Omega) \, \mathrm{d}r_m}{\int_{\mathfrak{S}_m} \hat{f}_{R_{(m)}}(\mathbf{r}_{(m-1)}, r_m; \Omega) \, \mathrm{d}r_m}, \quad (12)$$

where $\mathfrak{S}'_m = \mathfrak{S}_m \cap (-1, r_m]$ and \mathfrak{S}_m denotes the conditional support of R_m given $R_{(m-1)} = \mathbf{r}_{(m-1)}$. It is given by

$$\mathfrak{S}_m = \{r : -1 \leq r_{\min} < r < r_{\max} \leq 1\},$$

where r_{\min} and r_{\max} are such that, for r_m outside these values, $\mathbf{r}_{(m)}$ does not correspond to the ACF of a stationary AR(m) process. These values can be found as follows: Assume that $\mathbf{r}_{(m-1)}$ lies in the support of the distribution of $R_{(m-1)}$. From (2), the range of support for $R_{(m)}$ is determined by the inequality constraint $|\mathbf{R}_m| > 0$. Expressing the determinant in block form,

$$|\mathbf{R}_m| = |\mathbf{R}_{m-1}| \left(1 - \mathbf{r}'_{(m)} \mathbf{R}_{m-1}^{-1} \mathbf{r}_{(m)} \right),$$

and, as $|\mathbf{R}_{m-1}| > 0$ by assumption, r_m ranges over

$$\{r_m : 1 - \mathbf{r}'_{(m)} \mathbf{R}_{m-1}^{-1} \mathbf{r}_{(m)} > 0\}. \quad (13)$$

Letting $\mathbf{R}_{m-1}^{-1} = [w_{ij}]$, $i, j = 1, \dots, m$, and using the fact that \mathbf{R}_{m-1} is both symmetric and persymmetric (i.e., symmetric in the northeast-to-southwest diagonal), the range of r_m is given by

$$1 - \sum_{i=1}^m \sum_{j=1}^m w_{ij} r_i r_j = 1 - w_{mm} r_m^2 - 2r_m \sum_{i=1}^{m-1} w_{im} r_i - \sum_{i=1}^{m-1} \sum_{j=1}^{m-1} w_{ij} r_i r_j > 0.$$

This equation is quadratic in r_m so that the two ordered roots can be taken as the values for r_{\min} and r_{\max} respectively. For $m = 2$, this yields $2r_1^2 - 1 < r_2 < 1$; for $m = 3$,

$$\frac{(r_1 + r_2)^2}{r_1 + 1} - 1 < r_3 < \frac{(r_1 - r_2)^2}{r_1 - 1} + 1.$$

While computation of (12) is straightforward, it is preferable to derive an approximation to τ_m similar in spirit to the Lugannani and Rice (1980) saddlepoint approximation to the cdf of a univariate random variable. This method begins with the cumulative integral of the conditional saddlepoint density integrated over $(-1, r_m]$ as a portion of \mathfrak{S}_m , the support of R_m given $R_{(m-1)} = r_{(m-1)}$, as specified in the first equality of

$$\tau_m = \int_{\mathfrak{S}_m \cap (-1, r_m]} \hat{f}_{R_m | R_{(m-1)}}(r) \, dr = \int_{-\infty}^{w_0} h(w) \phi(w) \, dw. \quad (14)$$

A change of variable in the integration from dr to dw is needed which allows the integration to be rewritten as in the second equality of (14). Here, $\phi(w)$ is the standard normal density, $h(w)$ is the remaining portion of the integrand for equality to hold in (14), and w_0 is the image of r_m under the mapping $r \rightarrow w$ and given in (16) below. Temme's approximation approximates the right-hand-side of (14) and this leads to the expression

$$\Pr(R_m \leq r_m \mid R_{(m-1)} = r_{(m-1)}; \Omega) \simeq \Phi(w_0) + \phi(w_0) \left(\frac{1}{w_0} - \frac{1}{v_0} \right), \quad (15)$$

for $\hat{s}_m \neq 0$, where

$$w_0 = \text{sgn}(\hat{s}_m) \sqrt{\log(|\hat{\mathbf{P}}_m| / |\hat{\mathbf{P}}_{m-1}|)}, \quad (16)$$

$$v_0 = \hat{s}_m (|\hat{\mathbf{H}}_m| / |\hat{\mathbf{H}}_{m-1}|)^{\frac{1}{2}} \left[\text{tr}(\hat{\mathbf{P}}_{m-1}^{-1}) / \text{tr}(\hat{\mathbf{P}}_m^{-1}) \right]^{m-1}, \quad (17)$$

and Φ and ϕ denote the cdf and pdf of the standard normal distribution, respectively. Details of this derivation are given in Butler and Paoletta (1998) and Butler (2007, sects. 12.5.1 and 12.5.5).

In general, it is well known that the middle integral in (14) is less accurate than the normalized ratio in (12); see Butler (2007, eq. 2.4.6). However, quite remarkably, the Temme argument applied to (14) most often makes (15) more accurate than the normalized ratio in (12). A full and definitive answer to the latter tendency remains elusive. However, a partial asymptotic explanation is discussed in Butler (2007, sect. 2.3.2). In this setting, both (15) and (12) indeed yield very similar results, differing only in the third significant digit. It should be noted that the resulting p -values, or even the conditional distribution itself, cannot easily be obtained via simulation, so that it is difficult to check the accuracy of (12) and (15). In the next two sections, we show a way that lends support to the correctness (and usefulness) of the methods.

4. Properties of the Testing Procedure Under the Null

We first wish to assess the accuracy of the conditional saddlepoint approximation to (9) when $\Psi^{-1} = \mathbf{I}_T$ in (6), i.e., there is no autocorrelation in the disturbance terms of the linear model $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$. In this case, we expect $\tau_i \stackrel{\text{iid}}{\sim} \text{Unif}(0, 1)$. This was empirically tested by computing τ_1, τ_2, τ_3 and τ_4 in (10), based on (15), using observed values r_1, \dots, r_4 , for 10,000 replications of T -length time series, each consisting of T independent standard normal simulated random variables, $T = 15$ and $T = 30$, but with mean removal, i.e., taking $\mathbf{X} = \mathbf{1}$. Histograms of the resulting τ_i , as shown in Figure 1, are in agreement with the uniform assumption. Furthermore, the absolute sample correlations between each pair of the τ_i were all less than 0.02 for $T = 15$ and less than 0.013 for $T = 30$.

These results are in stark contrast to the empirical distribution of the “ t -statistics” and the associated p -values of the maximum likelihood estimates (MLEs). For 500 simulated series, the model $y_t = \mu + \varepsilon_t$, $\varepsilon_t = \phi_1 \varepsilon_{t-1} + \phi_2 \varepsilon_{t-2} + \phi_3 \varepsilon_{t-3} + u_t$, $u_t \stackrel{\text{iid}}{\sim} N(0, \sigma^2)$, $t = 1, \dots, T$, was estimated using exact maximum likelihood with approximate standard errors of the parameters obtained by numerically evaluating the Hessian at the MLE. Figure 2 shows the empirical distribution of the $\tau_i = F(t_i)$, where t_i is the ratio of the MLE of ϕ_i to its corresponding approximate standard error, $i = 1, 2, 3$, and $F(\cdot)$ refers to the cdf of the Student’s t distribution with $T - 4$ degrees of freedom.¹ The top two rows correspond to $T = 15$ and $T = 30$; while somewhat better for $T = 30$, it is clear that the usual distributional assumption on the MLE t -statistics does not hold. The last two rows correspond to $T = 100$ and $T = 200$, for which the asymptotic distribution is adequate.

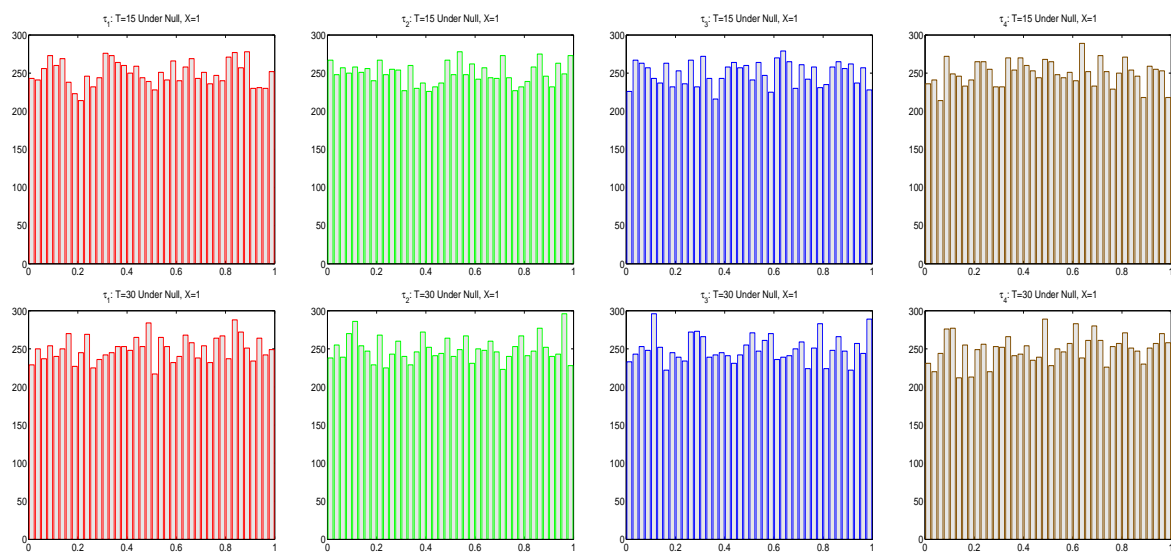


Figure 1. Histograms of τ_1, \dots, τ_4 , based on 10,000 replications with true data being iid normal, and taking $X = 1$. Top (bottom) panels are for $T = 15$ ($T = 30$).

¹ The ad hoc use of the Student’s t distribution was found to be slightly better than the standard normal for the smaller sample sizes.

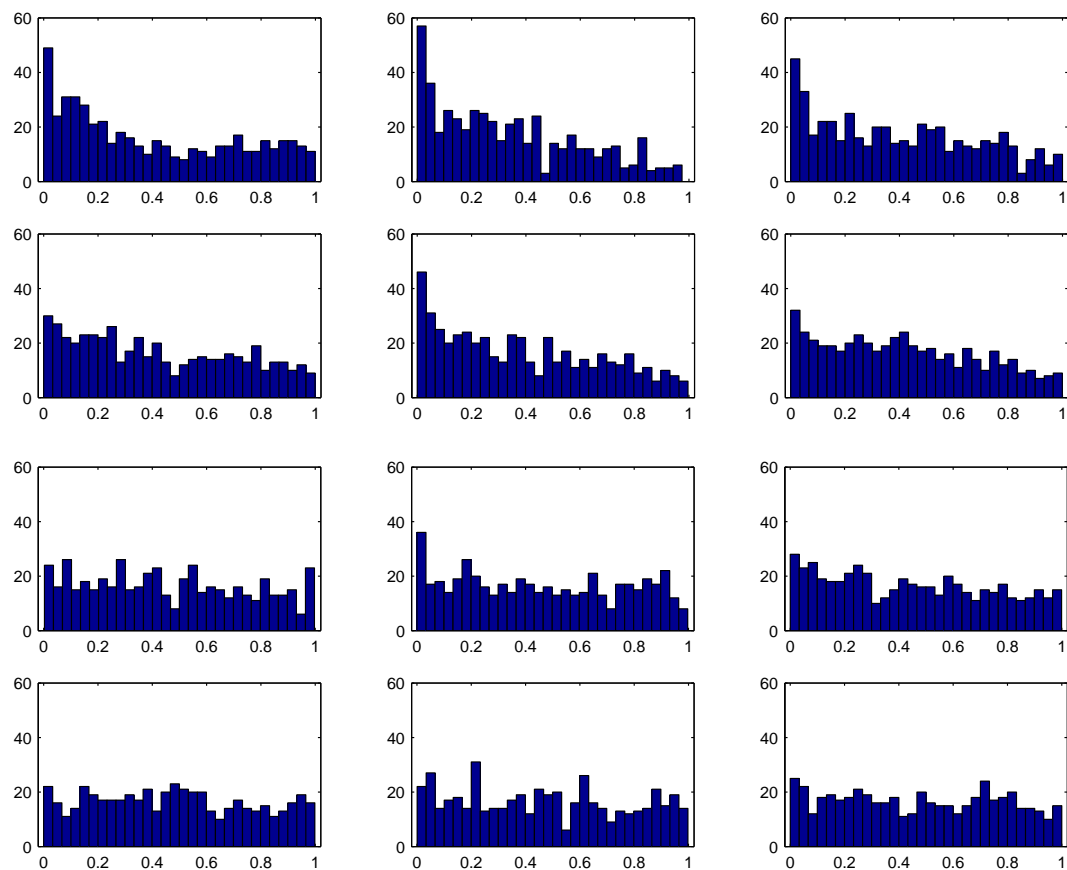


Figure 2. Empirical distribution of $F(t_i)$, $i = 1, 2, 3$ (left, middle and right panels) where t_i is the ratio of the MLE of ϕ_i to its corresponding approximate standard error and $F(\cdot)$ is the Student's t cdf with $T - 4$ degrees of freedom. Rows from top to bottom correspond to $T = 15$, $T = 30$, $T = 100$ and $T = 200$, respectively.

5. Properties of the Testing Procedure Under the Alternative

5.1. Implementation and Penalty-Based Criteria

One way of implementing the p -values computed from (15) for selecting the autoregressive lag order p is to let it be the largest value $j \in \{1, \dots, m\}$ such that $\tau_j < c$ or $\tau_j > 1 - c$; or set it to zero if no such extreme τ_j occurs. Hereafter, we refer to this as the *conditional ACF test*, or, in short, CACF. The effectiveness of this strategy will clearly be quite dependent on the choices of m and c . We will see that it is, unfortunately and like all selection criteria, also highly dependent on the actual, unknown autoregressive coefficients.

Another possible strategy, say, the *alternative CACF test*, is, for a given c and m , to start with testing an AR(1) specification and check if $\tau_1 < c$ or $\tau_1 > 1 - c$. If this is not the case, then one declares the lag order to be $p = 0$. If instead τ_1 is in the critical region, then one continues to the AR(2) case, inspecting if $\tau_2 < c$ or $\tau_2 > 1 - c$. If not, $p = 1$ is chosen; and if so, then τ_3 is computed, etc., continuing sequentially for $j = 1, \dots, m$, stopping when either the null at lag j is not rejected, in which case $p = j - 1$ is returned, or when $j = m$. Below, we will only investigate the performance of the first strategy. We note that the two strategies will have different small sample properties that clearly depend on the true p and the coefficients of the AR(p) model (as well as the sample size T and choices of m

and c). In particular, assuming in both cases that $m > p$, the alternative CACF test could perform better if m is chosen substantially larger than the true p .

While penalty function methods also require an upper limit m , the CACF has the extra “tuning parameter” c that can be seen as either a blessing or a curse. A natural value might be $c = 0.025$, so that, under the null of zero autocorrelation, p assumes a particular wrong value with approximate probability 0.05; and $p = 0$ is chosen approximately with probability $1 - 0.05m$, i.e.,

$$\begin{aligned}\Pr\{\text{choose } p = 0 \mid \text{white noise}\} &= \Pr\{\tau_j \in (0.025, 0.975), j = 1, \dots, m\} \\ &= (1 - 0.05)^m \approx 1 - 0.05m,\end{aligned}\quad (18)$$

from the binomial expansion. This value of c will be used for two of the three comparisons below, while the last one demonstrates that a higher value of c is advisable.

The CACF will be compared to the use of the following popular penalty-based measures, such that the lag order is chosen based upon the value of j for which they are minimized:

$$\begin{aligned}\text{AIC} &= \ln \hat{\sigma}^2 + \frac{2z}{T}, & \text{AICC} &= \ln \hat{\sigma}^2 + \frac{T+z}{T-z-2}, & \text{SBC} &= \ln \hat{\sigma}^2 + \frac{z \ln T}{T}, \\ \text{FPE} &= \hat{\sigma}^2 \frac{T+z}{T-z}, & \text{HQ} &= \ln \hat{\sigma}^2 + 2 \frac{z \ln \ln T}{T},\end{aligned}$$

where $\hat{\sigma}^2$ denotes the MLE of the innovation variance and $z = j + k$ denotes the number of estimated parameters (not including σ^2 , but including k , the number of regressors). Observe that both our CACF method, and the use of penalty criteria, assume the model regressor matrix is correctly specified, and condition on it, but both methods do not assume that the regression coefficients are known, and instead need to be estimated, along with the autoregressive parameters. Original references and ample discussion of these selection criteria can be found in the survey books of [Choi \(1992, chp. 3\)](#), [McQuarrie and Tsai \(1998\)](#) and [Konishi and Kitagawa \(2008\)](#), as well as [Lütkepohl \(2005\)](#) for the vector autoregressive case.

5.2. Comparison with AR(1) Models

For each method, the optimal AR lag orders among the choices $p = 0$ through $p = 4$ were determined for each of 100 simulated mean-zero AR(1) series of length $T = 30$ and AR parameter ϕ , using $\phi = 0, 0.1, 0.3, \dots, 0.9$,² as well as the two regression models, constant ($\mathbf{X} = \mathbf{1}$) and constant-trend ($\mathbf{X} = [\mathbf{1} \quad \mathbf{t}]$), where $\mathbf{t} = (1, 2, \dots, T)'$. For the CACF method, values $c = 0.025$ and $m = 4$ were used.

Figures 3 and 4 show the results for the two regression cases. In both cases, the CACF method dominates in the null model, while for small absolute values of ϕ , the CACF underselects more than the penalty-based criteria. For $\phi \geq 0.5$, the CACF dominates, though SBC is not far off in the $\mathbf{X} = \mathbf{1}$ case. For the constant-trend model, the CACF is clearly preferred. The relative improvement of the CACF performance in the constant-trend model shows the benefit of explicitly taking the regressors into account when computing tail probabilities in (15) via (3).

A potential concern with the CACF method is what happens if m is much larger than the true p . To investigate this, we stay with the AR(1) example, but consider only the case with $\phi = 0.5$, and use

² Negative values of ϕ were also considered; the results essentially paralleled their positive counterparts.

three choices of m , namely 2, 6 and 10. We first do this with a larger sample size of $T = 60$ and no \mathbf{X} matrix, which should convey an advantage to the penalty-based measures relative to CACF. For the former, we use only the AICC and BIC. Figure 5 shows the results, based on 1,000 replications. With $m = 2$, all three methods are very accurate, with CACF and BIC being about equal with respect to the probability of choosing the correct p of one, and slightly beating AICC. With $m = 6$, the BIC dominates. The nature of the CACF methodology is such that, when m is much larger than p , the probability of overfitting (choosing p too high) will increase, according to the choice of c . With $m = 10$, this is apparent. In this case, the BIC is superior, also substantially stronger than the more liberal AICC.

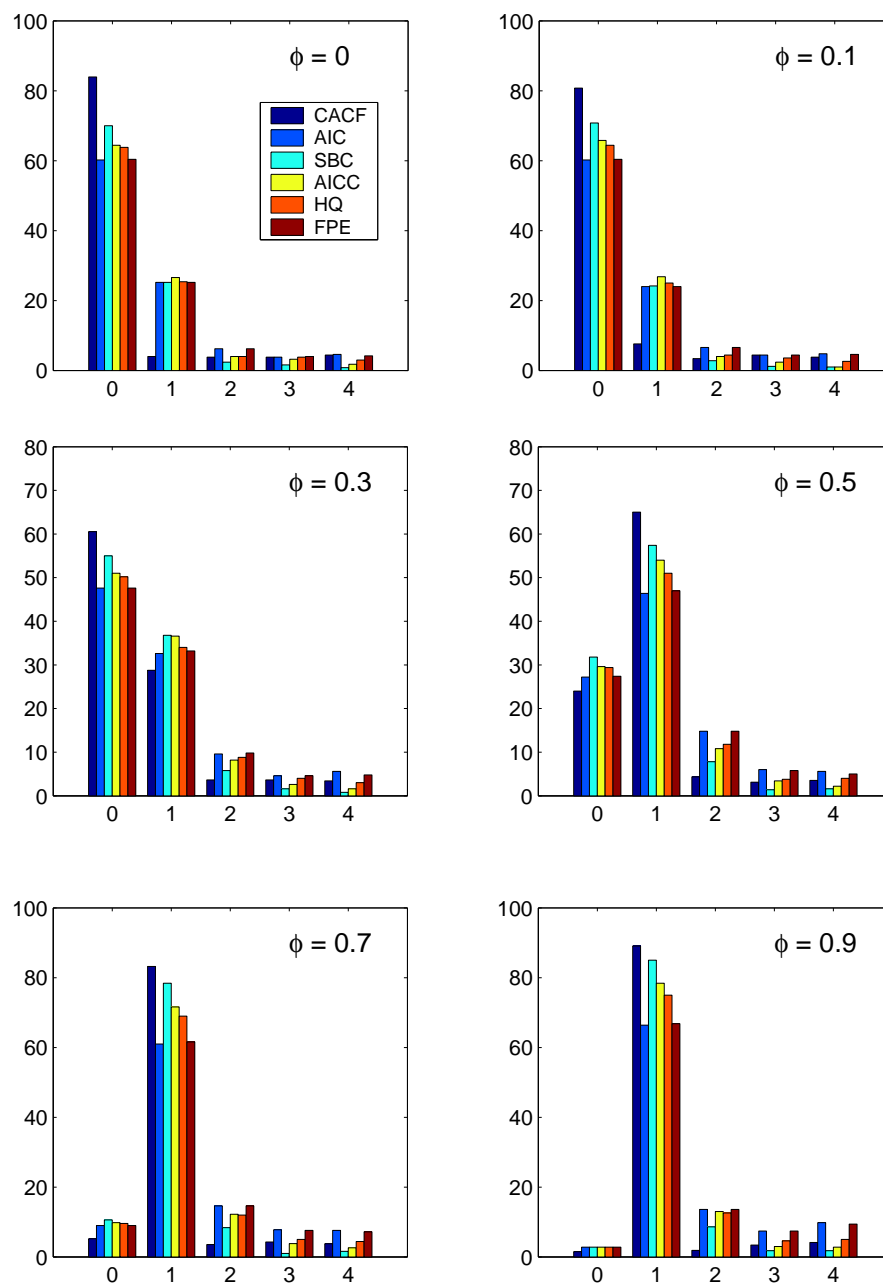


Figure 3. Performance of the various methods in the AR(1) case using $\mathbf{X} = \mathbf{1}$, $T = 30$, $m = 4$, and $c = 0.025$.

We now conduct a similar exercise, but using conditions for which the CACF method was designed, namely a smaller sample size of $T = 30$ and a more substantial regressor matrix of an intercept and time trend regression, i.e., $\mathbf{X} = [\mathbf{1}, \mathbf{t}]$. Figure 6 shows the results. For $m = 2$, the CACF clearly outperforms the penalty-based criteria, while for $m = 6$, which is substantially larger than the true $p = 1$, the CACF chooses the correct p with the highest probability of the three selection methods, though the AICC is very close. For the very large $m = 10$ (which, for $T = 30$, might be deemed inappropriate), CACF and BIC perform about the same with respect to the probability of choosing the correct p of one, while AICC dominates. Thus, in this somewhat extreme case (with $T = 30$ and $m = 10$), the CACF still performs competitively, due to its nearly exact small sample distribution theory and the presence of an \mathbf{X} matrix.

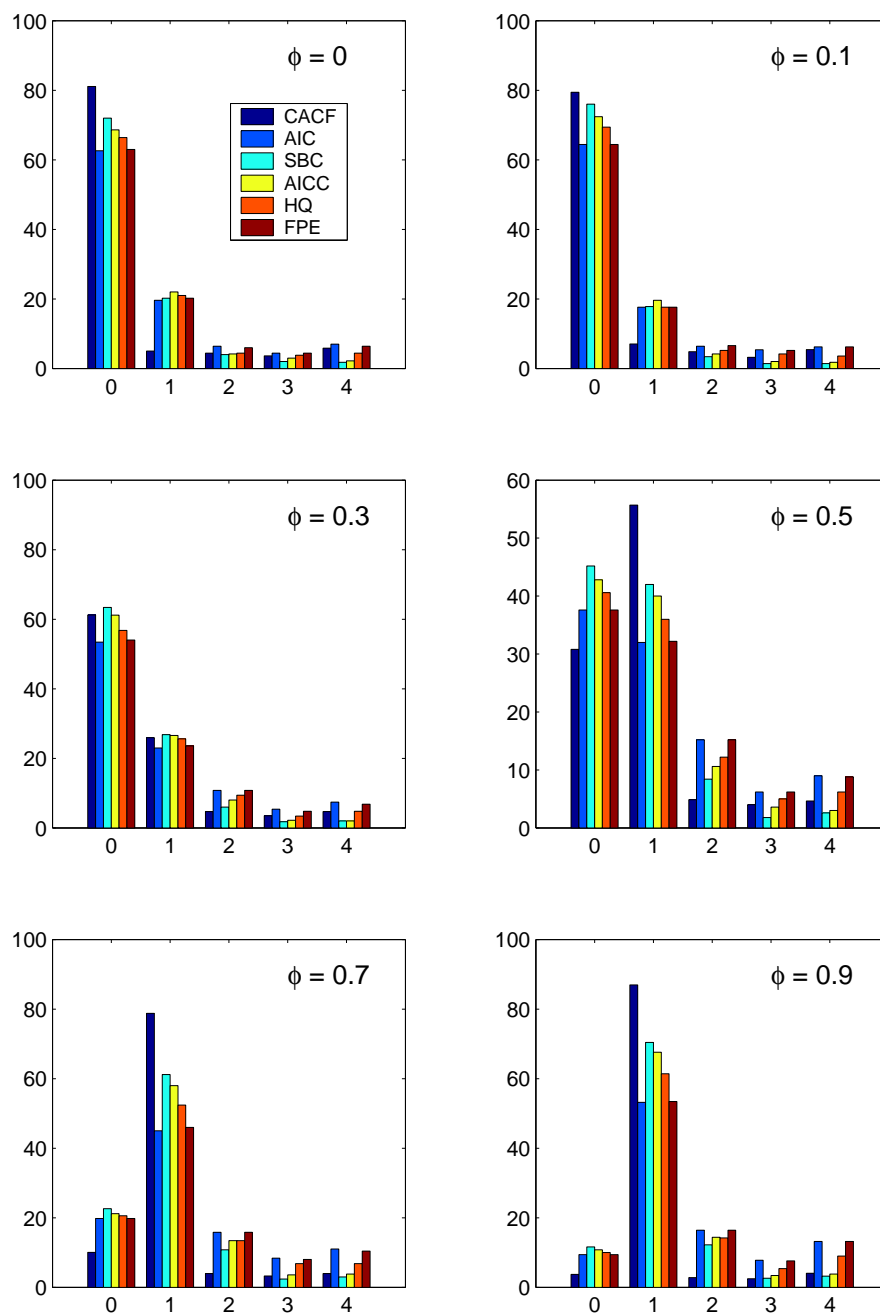


Figure 4. Similar to Figure 3, but based on $\mathbf{X} = [\mathbf{1}, \mathbf{t}]$.

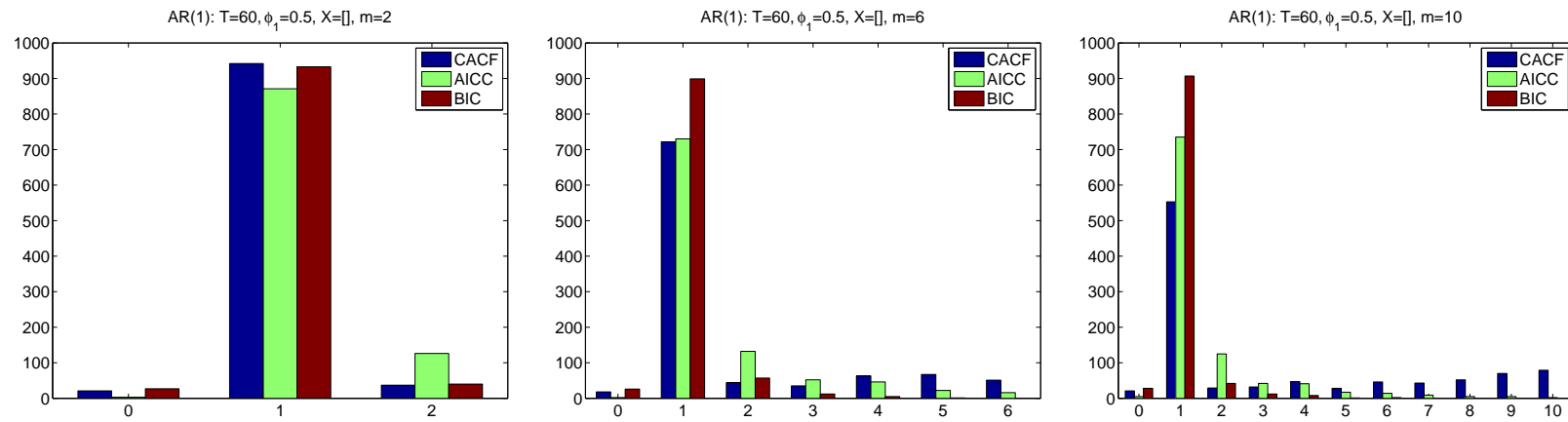


Figure 5. Histograms corresponding to the chosen value of AR lag length p , based on the CACF, AICC and BIC, using three values of tuning parameter m (2, 6 and 10, from left to right), and 1000 replications. True model is Gaussian AR(1) with parameter $\phi = 0.5$, sample size $T = 60$, and known mean (no X matrix). The CACF method uses $c = 0.025$.

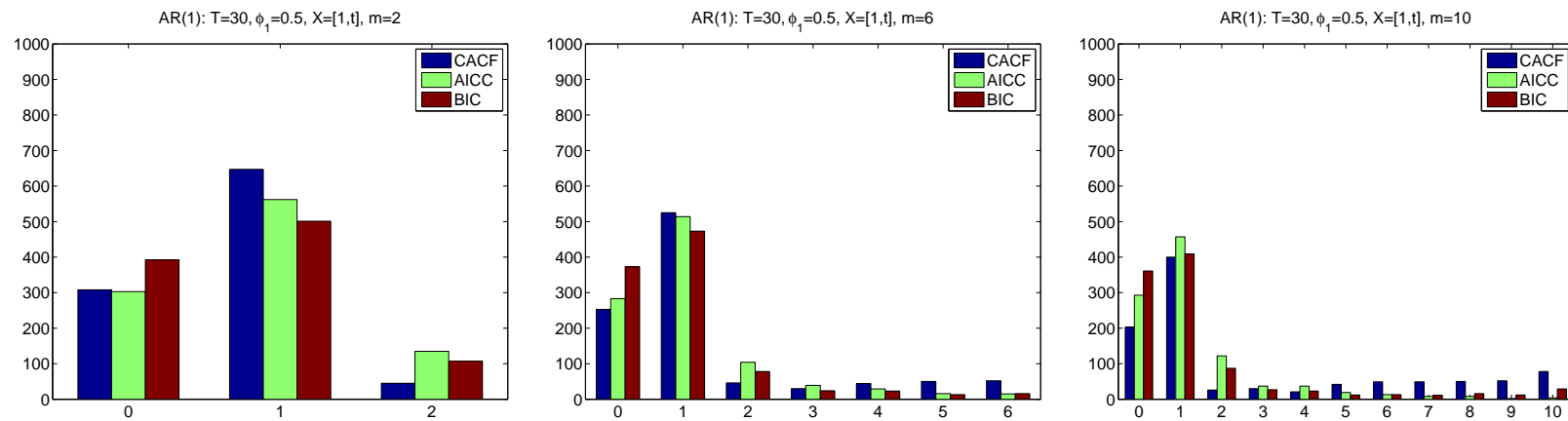


Figure 6. Similar to Figure 5, but for sample size $T = 30$ and $X = [1, t]$.

5.3. Comparison with AR(2) Models

The effectiveness of the simple lag order determination strategy is now investigated for the AR(2) model, but by applying it to cases for which the penalty-based criteria will have a comparative advantage, namely with more observations ($T = 60$) and either no regressors or just a constant. We also include the constant-trend case ($\mathbf{X} = [\mathbf{1}, \mathbf{t}]$) used above in the AR(1) simulation. Based on 1000 replications, six different AR(2) parameter constellations are considered. As before, the CACF method uses $c = 0.025$ and the highest lag order computed is $m = 4$.

The simulation results of the CACF test are summarized in Table 1. Some of these results are also given in the collection of more detailed tables in Appendix A. In particular, they are shown in the top left sections of Tables A1 (corresponds to no \mathbf{X} matrix) and A2 ($\mathbf{X} = \mathbf{1}$). There, the magnitude of the roots of the AR(2) polynomial are also shown, labeled as ζ_1 and ζ_2 , the latter being omitted to indicate complex conjugate roots (in which case $\zeta_1 = \zeta_2$).

The first model corresponds to the null case $p = 0$, in which the error rate for false selection of p should be (given the use of $c = 0.025$) about 5% for each false candidate. This is indeed the case, as the error rate with $m = 4$ should be roughly 0.80, as seen in the boldface entries. For the remaining non-null models, quite different lag selection characteristics were observed, with the choice of $p = 2$ ranging between 40% and 91% among the five AR(2) models in the known mean case; 33% and 91% in the constant but unknown ($\mathbf{X} = \mathbf{1}$) case; and 31% and 90% in the constant-trend ($\mathbf{X} = [\mathbf{1}, \mathbf{t}]$) case. Observe how, as expected from the small-sample theory, as the \mathbf{X} matrix increases in complexity, there is relatively little effect on the performance of the method, for a given AR(2) parametrization. However, the choice of the latter does have a very strong impact on its performance. For example, the 5th model is such that the CACF method chooses $p = 1$ more often than the correct $p = 2$.

Table 1. Simulation results of the CACF method for sample size $T = 60$: Percentages of the 1000 replications that chose AR lag length 0, 1, 2, 3, or 4, for six AR(2) models, with AR parameters a_1 and a_2 as indicated. The models, under the column denoted \mathcal{M} , are labeled as 1.0, etc., with the zero after the decimal point indicating a true MA(1) coefficient of $\theta = 0$. (Tables A1 and A2 show the results for nonzero θ_1 .) The panel denoted “No \mathbf{X} ” indicates the known mean case, while “ $\mathbf{X} = \mathbf{1}$ ” and “ $\mathbf{X} = [\mathbf{1}, \mathbf{t}]$ ” refers to the cases with unknown but constant mean, and constant and time trend, respectively. **Bold faced** numbers indicate the percentage of times the true AR lag order was correctly chosen.

\mathcal{M}	AR Param		No \mathbf{X}					$\mathbf{X} = \mathbf{1}$					$\mathbf{X} = [\mathbf{1}, \mathbf{t}]$				
			AR Lag Order (%)					AR Lag Order (%)					AR Lag Order (%)				
	a_1	a_2	0	1	2	3	4	0	1	2	3	4	0	1	2	3	4
1.0	0	0	81.5	4.5	4.3	5.2	4.5	80.3	3.9	5.0	6.0	4.8	82.0	3.1	5.3	4.9	4.7
2.0	1.2	−0.8	0.0	0.0	91.4	5.0	3.6	0.0	0.0	91.0	5.3	3.7	0.0	0.1	89.5	5.8	4.6
3.0	0.7	−0.3	0.1	35.2	56.0	4.2	4.5	0.0	35.9	54.4	4.8	4.9	0.1	37.7	54.0	3.9	4.3
4.0	0.4	−0.3	10.7	27.0	53.5	3.7	5.1	10.0	28.3	53.1	4.8	3.8	11.6	26.4	54.3	3.8	3.9
5.0	1.4	−0.45	0.0	52.5	40.2	2.6	4.7	0.0	59.9	33.0	2.9	4.2	0.0	57.6	31.3	5.6	5.5
6.0	−0.3	0.55	0.1	5.3	69.1	11.7	13.8	0.3	6.5	66.7	12.5	14.0	0.0	8.1	66.1	10.8	15.0

The comparative results using the penalty-based methods are shown in Tables A3 through A7 under the heading “For AR(p) Models”. We will discuss only the $\mathbf{X} = \mathbf{1}$ case in detail. For the null model (i.e., $p = 0$), denoted 1.0x, where the “x” indicates that an \mathbf{X} matrix was used, the CACF outperforms all other criteria by a wide margin, with 80.3% of the runs resulting in $p = 0$, compared with the 2nd best, SBC, with 66.2%. For model 2.0x, all the model selection criteria performed well, with the CACF and SBC resulting in 91.0% and 91.6%, respectively. Similarly, all criteria performed relatively poorly for models 3.0x and 4.0x, but particularly the CACF, which was worst (with $p = 2$ occurring 54.4% and 53.1% of the time, respectively), while AICC and HQ were the best and resulted in virtually the same $p = 2$ values for the two models (70% and 65%). Similar results hold for models 5.0x and 6.0x, for which the CACF again performs relatively poorly.

The unfortunate and, perhaps unexpected, fact that the performance of the new method and the penalty-based criteria highly depend on the true model parameters is not new; see, for example, Rahman and King (1999) and the references therein.

5.4. Higher Order AR(p) Processes

Clearly, as p increases, it becomes difficult to cover the whole parameter space in a simulation study. We (rather arbitrarily) consider AR(4) models with parameters $\phi_1 = 0.4$, $\phi_2 = -0.3$, $\phi_3 = 0.2$ and ϕ_4 takes on the 6 values -0.1 through -0.6 , for which the maximum moduli of the roots of the AR polynomial are 0.616, 0.679, 0.747, 0.812, 0.864 and 0.908, respectively. This is done for two sample sizes, $T = 30$ and $T = 60$ and, in an attempt to use a more complicated design matrix that is typical in econometric applications, an \mathbf{X} matrix corresponding to an intercept-trend model with structural break, i.e., for $T = 30$,

$$\mathbf{X}' = \begin{bmatrix} 1 & 1 & 1 & \cdots & 1 & \cdots & 1 \\ 1 & 2 & 3 & \cdots & 16 & \cdots & 30 \\ 0 & 0 & 0 & \cdots & 1 & \cdots & 1 \\ 0 & 0 & 0 & \cdots & 1 & \cdots & 15 \end{bmatrix}.$$

As mentioned in the beginning of Section 5, the choices of c and m are not obvious, and ideally would be purely data driven. The optimal value of m in conjunction with penalty-based criteria is still an open issue; see, for example, the references in Choi (1992, p. 72) to work by E. J. Hannan and colleagues. The derivation of a theoretically based optimal choice of m for the CACF is particularly difficult because the usual appeal to asymptotic arguments is virtually irrelevant in this small-sample setting. At this point, we have little basis for their choices, except to say (precisely as others have) that m should grow with the sample size.

It is not clear if the optimal value of c should vary with sample size. What we can say is that its choice depends on the purpose of the analysis. For example, consider the findings of Fomby and Guilkey (1978), who demonstrated that, when measuring the performance of $\hat{\beta}$ based on mean squared error, the optimal size of the Durbin-Watson test when used in conjunction with a pretest estimator for the AR(1) term should be much higher than the usual 5%, with 50% being their overall recommendation. This will, of course, increase the risk of model over-fitting, but that can be somewhat controlled in this context by a corresponding reduction in m . (Recall that the probability of not rejecting the null hypothesis of no autocorrelation is $100(1 - 2cm)\%$ under white noise.)

For the trials in the AR(4) case, we take $m = 7$ to provide some room for over-fitting (but which admittedly might be considered somewhat high for only $T = 30$ observations). With this m , use of $c = 0.125$ proved to be a good choice for all the runs with $T = 30$. For this sample size, the top panels of Figures 7 and 8 show the outcomes for all criteria for each of the six values of ϕ_4 . For all the criteria, the probability of choosing $p = 4$ increases as ϕ_4 increases in magnitude.

To assist the interpretation of these plots, we computed the simple measures $L_1 = \sum |\hat{p}_i - 4|$ and $L_2 = \sum (\hat{p}_i - 4)^2$ for each criteria. In all six cases, the CACF was the best according to both L_1 and L_2 . The lower panels of Figures 7 and 8 show just the CACF and the criteria that was second best according to L_2 . For $\phi_4 = -0.1, \dots, -0.4$, the FPE was the second best, while for $\phi_4 = -0.5$ and -0.6 , HQ and AICC were second best, respectively.

This is also brought out in Table 2, which shows the L_2 measures for the two extreme cases $\phi_4 = -0.1$ and -0.6 . Observe how, in the former, the SBC and AICC perform significantly worse than the other criteria, while for $\phi_4 = -0.6$, they are 2nd and 3rd best. This example clearly demonstrates the utility of the new CACF method, and also emphasizes how the performance of the various penalty-based criteria are highly dependent on the true autoregressive model parameters.

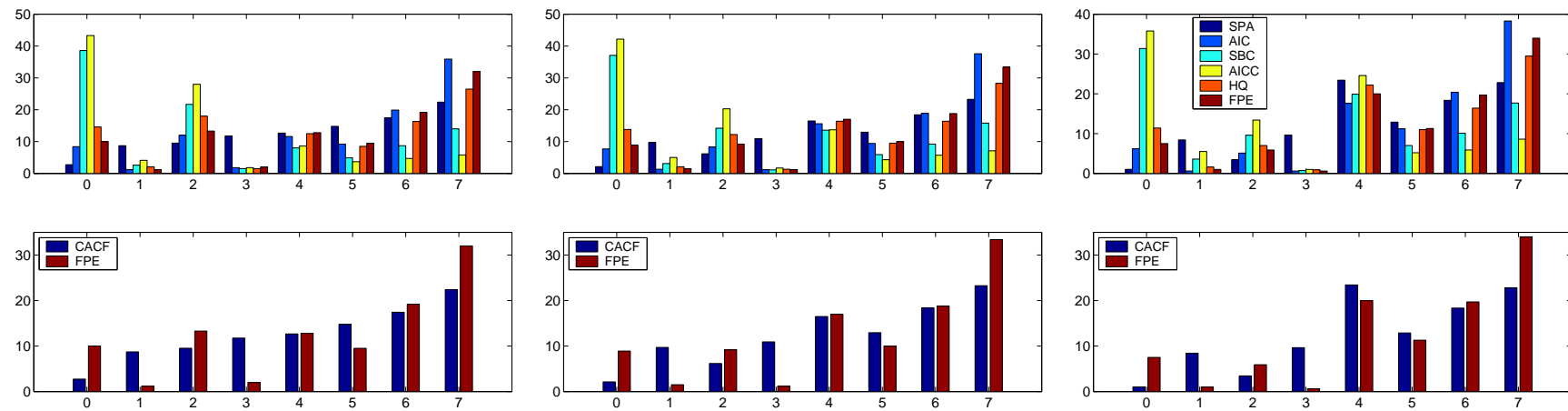


Figure 7. For AR(4) process with $T = 30$ and $\phi_4 = -0.1$ (left), -0.2 (middle) and -0.3 (right). Top panels show all criteria, with SPA = CACF. Bottom panels show just the CACF and the 1st or 2nd best penalty-based criteria: If the CACF was best, it is given first in the legend, otherwise second.

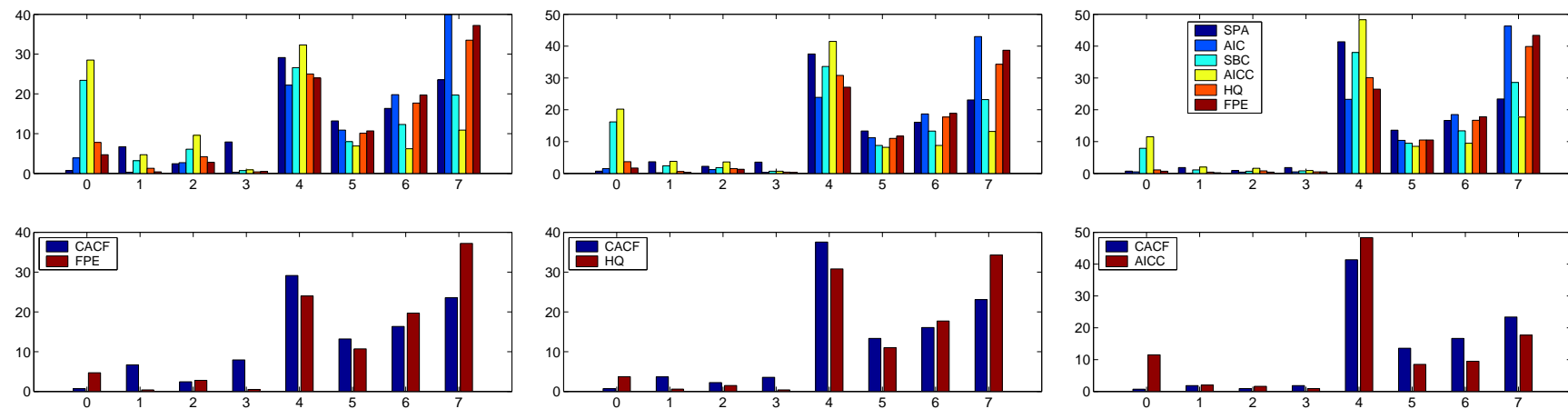


Figure 8. Same as Figure 7 but for $\phi_4 = -0.4$ (left), -0.5 (middle) and -0.6 (right).

Table 2. Summary measures for the two extreme AR(4) models. The measure $L_2 = \sum(\hat{p}_i - 4)^2$ and L_2 scaled by its minimum observed value, for the two model cases $\phi_4 = -0.1$ and -0.6 .

$\phi_4 = -0.1$	CACF	FPE	AIC	HQ	SBC	AICC
L_2	458	600	607	638	895	918
$L_2/458$	1.000	1.311	1.325	1.393	1.954	2.005
$\phi_4 = -0.6$	CACF	AICC	SBC	HQ	FPE	AIC
L_2	324	415	460	461	487	512
$L_2/324$	1.000	1.280	1.420	1.423	1.503	1.580

The results for the $T = 60$ case were not as unambiguous. For each of the six models, a value of c could be found that rendered the CACF method either the best or a close second best. These ranged from $c = 0.125$ for $\phi_4 = -0.1$ to $c = 0.01$ for $\phi_4 = -0.6$. When using a constant value of c between these two extremes such as $c = 0.05$, the AIC and FPE performed better for $\phi_4 = -0.1$, while SBC and AICC were better for $\phi_4 = -0.6$.

5.5. Parsimonious Model Selection

The frequentist order selection methods such as the the sequential one proposed herein and—even more so—the penalty-based methods, have the initial appeal of being objective. However, the infusion of “prior information”, most often in the forms of non-quantitative beliefs and opinions, is ubiquitous when building models with real data; the well-known quote from Edward Leamer (“There are two things you are better off not watching in the making: sausages and econometric estimates”) immediately comes to mind. Indeed, the empirical case studies by Pankratz (1983) make clear the desire for parsimonious models and attempt to “downplay” the significance of spikes in the correlograms or p -values of estimates at higher order lags that are not deemed “sensible”. Moreover, any experienced modeler has a sense, explicit or not, of the maximally acceptable number of model parameters allowed with respect to the available sample size, and is (usually) well aware of the dangers of overfitting.³

Letting the level of a test decrease with increasing sample size is common and good practice in many testing situations; for discussion see Lehmann (1958), Sanathanan (1974), and Lehmann (1986, p. 70). When the level of the test is fixed and the sample size is allowed to increase, the power can increase to unacceptably high values, forcing the type I and type II errors to be out of balance. Indeed, such balance is important in the model selection context, because the consequences of the two error types are not especially different as they would be in the more traditional testing problem. Furthermore, there may be a desire to have the relative values of the two error types reflect the desire for parsimony. In the context of our sequential test procedure, such preferences would suggest using smaller levels when testing for larger lag m , and increasingly larger levels as the lag order tested becomes smaller. Thus, at each stage of the testing, balance may be achieved between the two types of errors to reflect the desired parsimony.

This frequentist “trap” could be avoided by conducting an explicit Bayesian analysis instead. While genuine Bayesian methods have been pursued (see, for example, Zellner 1971; Schervish and Tsay 1988; Chib 1993; Le et al. 1996; Brooks et al. 2003; and the references therein), their popularity is quite limited. The CACF method proposed herein provides a straightforward method of incorporating the preference for low order models, while still being as “objective” as any frequentist hypothesis testing procedure can, for which the size of the test needs to be chosen by the analyst. This is achieved by letting the tuning parameter c be a vector with different values for different lags. For example, with the AR(4)

³ The ill effects of data dredging and autoregressive model over-fitting are well discussed in Chatfield (2001, chp. 8) and the references therein.

case considered previously, the optimal choice would be $\mathbf{c} = [0.5, 0.5, 0.5, 0.5, 0, 0, 0]$. While certainly such flexibility could be abused to arrive at virtually any desired model, it makes sense to let the elements of \mathbf{c} decrease in a certain manner if there is a preference for low-order parsimonious models. Notice also that, if seasonal effects are expected, then the \mathbf{c} vector could also be chosen to reflect this.

To illustrate, we applied the simple linear sequence

$$\mathbf{c} = [0.175, 0.15, 0.125, 0.10, 0.075, 0.05, 0.025] \quad (19)$$

to the $T = 60$ case. This resulted in the CACF method being the best for all values of ϕ_4 except the last, $\phi_4 = -0.6$, in which case it was slightly outperformed by the SBC. As was expected, when using (19) for the time series with $T = 30$, the CACF remained superior in all cases but by an even larger margin.

6. Mixed ARMA Models

Virtually undisputed in time series modeling is the sizeable increase in difficulty for identifying the orders in the mixed, i.e., ARMA, case. It is interesting to note that, in the multivariate time series case, the theoretical and applied literature is dominated by strict AR models (vector autoregressions), whether for prediction purposes or for causality testing. Under the assumption that the true data generating process does not actually belong to the ARMA class, it can be argued that, for forecasting purposes, purely autoregressive structures will often be adequate, if not preferred; see also Zellner (2001) and the references therein. Along these lines, it is also noteworthy that the thorough book on regression and time series model selection by McQuarrie and Tsai (1998) only considers order selection for autoregressive models in *both* the univariate and multivariate case. Their only use for an MA(1) model is to demonstrate AR lag selection in misspecified models.

Nevertheless, it is of interest to know how the CACF method performs in the presence of mixed models. While the use of penalty-based criteria for mixed model selection is straightforward, it is not readily apparent how the CACF method can be extended to allow for moving average structures. This section presents a way of proceeding.

6.1. Known Moving Average Structure

The middle left and lower left panels of Tables A1 and A2 show the CACF results when using the same AR structure as previously, but with a *known* moving average structure, i.e., (15) is calculated taking Ψ^{-1} corresponding to an MA(1) model. As with all model results in Tables A1 and A2, the sample size is $T = 60$. This restrictive assumption allows a clearer comparison of methods without the burden of MA order selection or estimation; this will be relaxed in the next section, so that the added effect of MA parameter estimation can be better seen.

Models 1.5 through 6.5 take the MA parameter θ to be 0.5, while models 1.9 through 6.9 use $\theta = 0.9$. For the null models 1.5 and 1.9, only a slight degradation of performance is seen. For the other models, on average, the probability of selecting $p > 2$ increases markedly as θ increases, drastically so for model 6.

The results can be compared to the entries in Tables A3 through A7 labeled “For ARMA($p, 1$) Models”, for which the penalty-based criteria were computed for the five models ARMA($i, 1$), $i = 0, \dots, 4$. (Notice that the comparison is not entirely fair because the CACF method has the benefit of knowing the MA polynomial, not just its length.) Consider the $\theta = 0.5$ cases first. For the null model 1.5x, the CACF performed as expected under the null hypothesis, with 78.8% choices of $p = 0$ and about 5% for each of the other choices. It was outperformed however by the SBC, with 86% $p = 0$ choices and an ever diminishing probability as p increases, which agrees with the SBC’s known properties of low model order preference. The $p = 0$ choices for the other criteria were between 54.2% (AIC and FPE) and 70.6% (HQ). For model 2.5x, in terms of number of $p = 2$ choices, CACF, AIC and FPE were virtually tied with about 63%, while SBC, HQ and AICC gave 87%, 77.4% and 71.6%, respectively. The CACF for models 3.5x and 4.5x performed—somewhat unexpectedly in

light of previous results—relatively well: For model 3.5x, CACF resulted in 61.2% choices of $p = 2$, while the SBC was the worst, with 37.4%. The others were all about 43%. Similarly for model 4.5x: CACF (56.8%), AIC (38.2%), SBC (25.8%), AICC (37.8%), HQ (33.8%) and FPE (38.2%). Model 5.5x resulted in (approximately) CACF, AIC and FPE (50%), and SBC, AICC and HQ (60.0%). For model 6.5x, the CACF was disastrous, with only 25.8% for $p = 2$, while the other criteria ranged from 71.2% (SBC) to 59.2% (AIC).

As before, the performance of all the criteria is highly dependent on the true model parameters. However, what becomes apparent from this study is the great disparity in performance: For a particular model, the CACF may rank best and SBC the worst, while for different parameter constellation, precisely the opposite may be the case. That the SBC is often among the best performers agrees with the findings of [Koreisha and Yoshimoto \(1991\)](#) and the references therein.

The results for the $\theta = 0.9$ cases do not differ remarkably from the $\theta = 0.5$ cases, except that for models 3.9x and 4.9x, all the penalty-based criteria were much closer to (and occasionally slightly better) than the CACF.

6.2. Iterative Scheme for ARMA(p, q) Models with q Known

Assume, somewhat more realistically than before, that the regression error terms follow a stationary ARMA(p, q) process with q known, but the actual MA parameters $\theta = (\theta_1, \dots, \theta_q)$ and p are not known. A possible method for eliciting p using the sequential test (15) is as follows: Iterate the two following steps starting with $p_1 = 0$: (i) Estimate an ARMA(p_i, q) model to obtain $\hat{\theta}$; (ii) Compute τ_1, \dots, τ_m with Ψ^{-1} corresponding to the MA(q) model with parameters $\hat{\theta}$, from which p_{i+1} is determined. Iteration stops when $p_{i+1} = p_i$ (or i exceeds some preset value, I), and p_{i+1} is set as before for a given value of c . The choice of m and c will clearly be critical to the performance of this method; simulation, as detailed next, will be necessary to determine its usefulness.

The right panels of Tables A1 and A2 show the results when applied to the same 500 simulated time series of length $T = 60$ as previously used, but with the aforementioned iterative scheme applied with $m = 4$ and $c = 0.025$ (also as before), $I = 5$ and $q = 1$. Unexpectedly, for the null model 1, $p = 0$ was actually chosen more frequently than under the known θ cases, for each choice of θ . Not surprisingly, for the $\theta = 0$ cases, the iterative schemes resulted in poorer performance compared to their known θ counterparts, with model under-selection (i.e., $p < 2$) occurring more frequently, drastically so for models 3 and 4. For $\theta = 0.5$, models 3 and 4 again suffer from under-selection, though less so than with $\theta = 0$, while the remaining models exhibit an overall mild improvement in lag order selection. In the $\theta = 0.9$ case, performance is about the same whether θ is known or not for all 6 models, though for model 6, high over-selection in the θ known case is reversed to high under-selection for θ not known.

The number of iterations required until convergence and the probability of not converging (given in the column labeled 6+) also depends highly on the true model parameters; models for which the true AR polynomial roots were smallest exhibited the fastest convergence. Nevertheless, in most cases, one or two iterations were enough, and the probability of non-convergence appears quite low. (The few cases that did not converge were discarded from the analysis.) While certainly undesirable, if the iterative scheme does not converge, it does not mean that the results are not useful. Most often, the iterations bounced back and forth between two choices, from which a decision could be made based on “the subjective desire for parsimony” and/or inspection of the actual p -values, which could be very close to the cutoff values, themselves having been arbitrarily chosen.

To more fairly compare the CACF results with the penalty-based criteria, the latter were evaluated using the 10 models AR(i), ARMA($i, 1$), $i = 0, \dots, 4$, the results of which are in Tables A3 through A7 under the heading “Among both Sets”. To keep the analysis short, Table 3 presents only the percentage of correct p choices for two criteria (CACF and the 1st or 2nd best). The CACF was best 6 times, SBC and AICC were each best 5 times, while HQ and FPE were best one time each. It must be emphasized that these numbers are a very rough reduction of the performance data. For example,

Table 3 shows that cases 2.0x and 2.5x were extremely close, while 6.0x, 3.5x, 4.5x, 5.5x, 1.9x were reasonably close. A fair summary appears to be that

- In most cases, either the CACF, SBC and AICC will be the best,
- Each of these can perform (sometimes considerably) better than the other two for certain parameter constellations,
- Each can perform relatively poorly for certain parameter constellations.

Table 3. Performance Summary. Selected information from Table A2 and Tables A3 through A7. For each model (with $\mathbf{X} = \mathbf{1}$), the percentage of correct p choices (0 for models 1.0x, 1.5x, 1.9x and 2 for the rest), are shown for the CACF and the penalty-based criteria that was either 1st or 2nd best. Form is criteria:percentage, where criteria 0 is the CACF, 1 is AIC, 2 is SBC, 3 is AICC, 4 is HQ and 5 is FPE. The criteria with the larger percentage is given first.

Model	1.0x		2.0x		3.0x		4.0x		5.0x		6.0x	
	0:87	2:79	0:88	2:87	3:47	0:14	3:40	0:16	2:65	0:53	0:79	2:74
Model	1.5x		2.5x		3.5x		4.5x		5.5x		6.5x	
	0:91	2:77	4:65	0:64	0:49	2:43	3:36	0:32	0:61	2:57	2:59	0:36
Model	1.9x		2.9x		3.9x		4.9x		5.9x		6.9x	
	2:86	0:83	2:87	0:59	3:53	0:42	3:52	0:33	2:74	0:60	5:22	0:06

7. Performance in the Non-Gaussian Setting

All our findings above are based on the assumptions that the true model is known to be a linear regression with correctly specified \mathbf{X} matrix; error terms are from a Gaussian $\text{AR}(p)$ process; tuning parameter m is chosen such that $m \geq p$; and parameters $p, a_1, \dots, a_p, \beta$, and σ^2 are fixed but unknown. We now modify this by assuming, similarly, that the true data generating process is $\mathbf{y} = \mathbf{X}\beta + \epsilon$, with $\epsilon_t = \phi\epsilon_{t-1} + U_t$ a stationary $\text{AR}(1)$ process, but now such that $U_t \stackrel{\text{iid}}{\sim} t_\nu(0, \sigma)$, $t = 1, \dots, T$, i.e., Student's t with ν degrees of freedom, location zero, and scale $\sigma > 0$.

Figures 9 and 10 depict the results for the CACF, AICC, and SBC methods, having used a grid of ϕ -values, true $p = 1$, a sample size of $T = 50$, $m = 5$, $c = 0.025$, and four different values of degrees of freedom parameter ν ; and for all methods falsely assuming Gaussianity. Figure 9 is for the known mean case, while Figure 10 assumes the constant and time-trend model $\mathbf{X} = [\mathbf{1}, \mathbf{t}]$.

We see that none of the methods are substantially affected by use of even very heavy-tailed innovation sequences, notably the CACF, which explicitly uses the normality assumption in the small-sample distribution theory. However, note from the top left panel of Figure 9 (which corresponds to $\nu = 1$, or Cauchy innovations) that, for $p = \phi = 0$, instead of $0.774 = (1 - 0.05)^m \approx 1 - 0.05m = 0.75$ from (18), the null of $p = 0$ is chosen about 86% of the time, while from the third and fourth rows (for $\nu = 5$ and $\nu = 200$), it is about 78%. This was to be expected: In the non-Gaussian case, (18) is no longer tenable, as (i) the τ_j are no longer independent; and (ii) their marginal distributions under the null will no longer be precisely (at least up to the accuracy allowed for by the saddlepoint approximation) $\text{Unif}(0, 1)$. In particular, the latter violation is such that the empirically obtained quantiles of τ_j under the null no longer match their theoretical ones, and, as seen, the probability that τ_j falls outside the range, say, $(0.025, 0.975)$, is smaller than 0.05. This results in the probability of choosing $p = 0$ when it is true being larger than the nominal of 0.774.

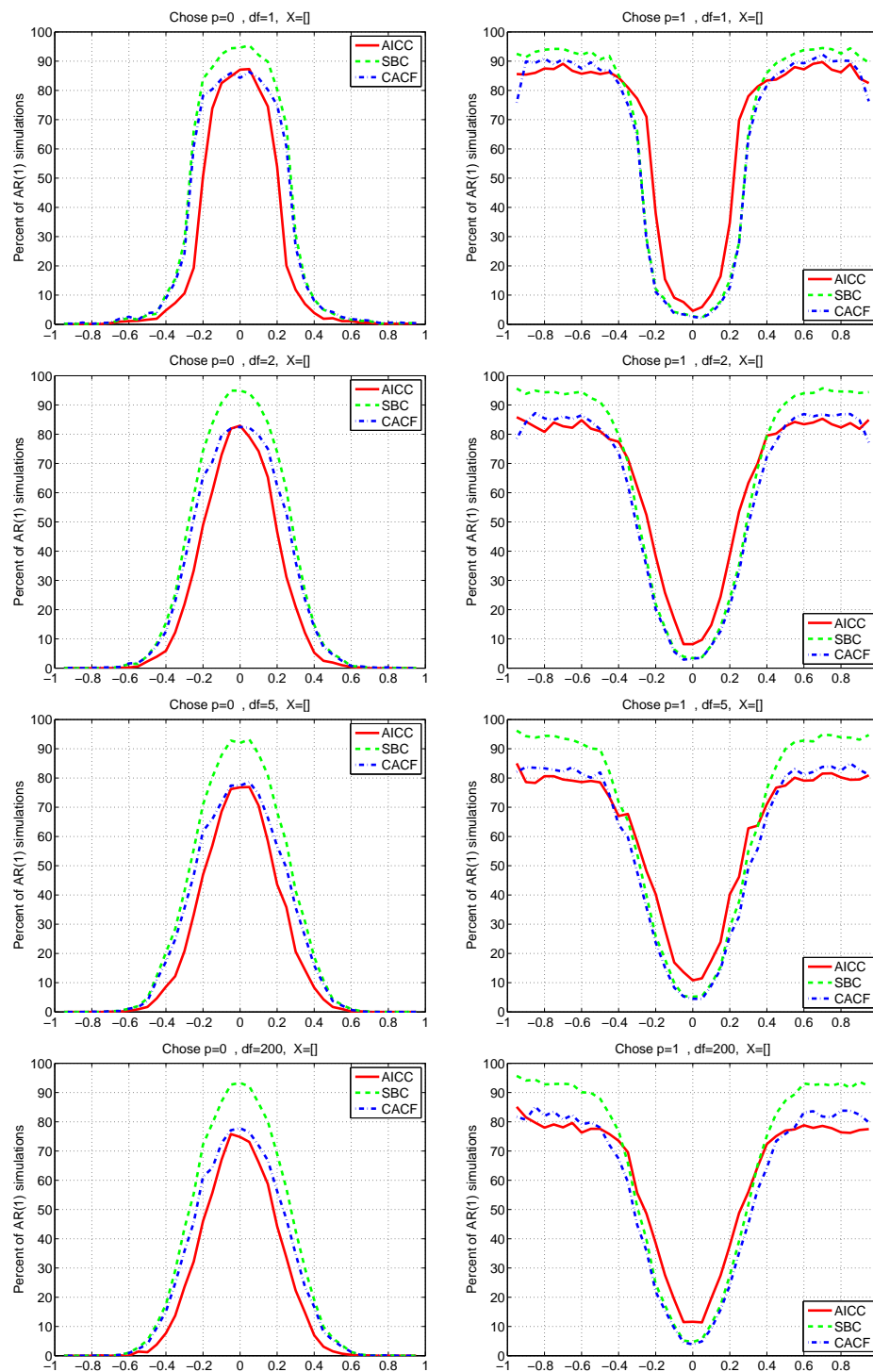


Figure 9. Performance of the three indicated AR order selection methods as a function of autoregressive parameter ϕ , for sample size $T = 50$, known mean (denoted by $X = []$) and $m = 5$, when the true AR order is $p = 1$ and (falsely) assuming Gaussianity. The true innovation sequence consists of i.i.d. Student's $t(\nu)$ realizations, with $df = \nu$ indicated in the titles (from top to bottom, $\nu = 1$, $\nu = 2$, $\nu = 5$, and $\nu = 200$). **Left (right)** panels indicate the percentage of the 1000 replications that resulted in choosing $p = 0$ ($p = 1$).

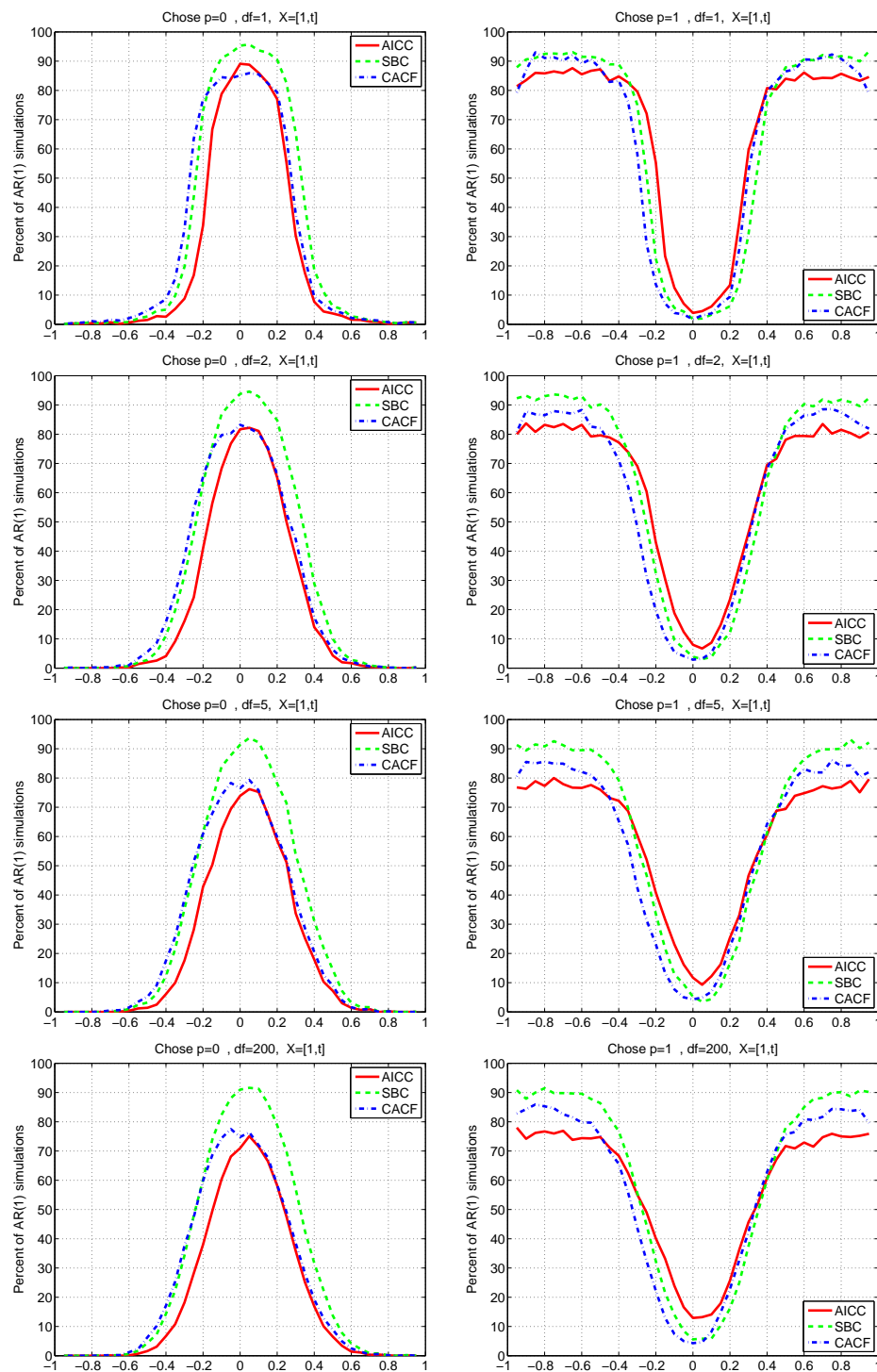


Figure 10. Same as Figure 9 but having used $X = [1, t]$.

Similarly, the choice of $p = 1$ when $p = \phi = 0$ should occur about 5% of the time for the CACF method, but, from the right panels of Figure 9, it is lower than this, decreasing as ν decreases. However, for $\nu = 5$, it is already very close to the nominal of 5%. Interestingly, with respect to choosing $p = 0$, the behavior of the CACF for all choices of ν is virtually identical to the AICC near $\phi = 0$, while as $|\phi|$ grows, the behavior of the CACF coincides with that of the SBC. (Note that this behavior is precisely what we do *not* want: Ideally, for $\phi = 0$, the method would always choose $p = 0$, while for $\phi \neq 0$, the method would never choose $p = 0$.)

Figure 10 is similar to Figure 9, but having used $\mathbf{X} = [\mathbf{1}, \mathbf{t}]$. Observe how, for all values of ν , unlike the known mean case, the performance of the AICC and SBC is no longer symmetric about $\phi = 0$, but the CACF is still virtually symmetric. Fascinatingly, we see from the left panels of Figure 10 that the CACF probability of choosing $p = 0$ virtually coincides with that of the AICC for $\phi \geq 0$, while for $\phi < -0.2$, it virtually coincides with that of the SBC.

8. Conclusions

We have operationalized the UMPU test for sequential lag order selection in regression models with autoregressive disturbances. This is made possible by using a saddlepoint approximation to the joint density of the sample autocorrelation function based on ordinary least squares residuals from arbitrary exogenous regressors and with an arbitrary (covariance stationary) variance-covariance matrix. Simulation results verify that, compared to the popular penalty-based model selection methods, the new method fairs very well precisely in situations that are both difficult and common in practice: when faced with small samples and an unknown mean term. With respect to the mean term, the superiority of the new method increases as the complexity of the exogenous regressor matrix \mathbf{X} increases; this is because the saddlepoint approximation explicitly incorporates the \mathbf{X} matrix, differing from the approximation developed in Durbin (1980), which only takes account of its size, or the standard asymptotic results for the sample ACF and sample partial ACF, which completely ignore the regressor matrix.

The simulation study also verifies a known (but—we believe—not *well-known*) fact that the small sample performances of penalty-based criteria such as SBC and (corrected) AIC are highly dependent on the actual autoregressive model parameters. The same result was found to hold true for the new CACF method as well. Autoregressive parameter constellations were found for which CACF was greatly superior to all other methods considered, but also for which CACF ranked among the worst performers. Based on the use of a wide variety of parameter sets, we conclude that the new CACF method, the SBC and the corrected AIC, in that order, are the preferred methods, although, as mentioned, their comparative performance is highly dependent on the true model parameters.

An aspect of the new CACF method that greatly enhances its ability and is not applicable with penalty-based model selection methods is the use of different sizes for the sequential tests. This allows an objective way of incorporating prior notions of preferring low order, parsimoniously parameterized models. This was demonstrated using a linear regression model with AR(4) disturbance terms; the results highly favor the use of the new method in conjunction with a simple, arbitrarily chosen, linear sequence of size values. More research should be conducted into finding optimal, sample-size driven choices of this sequence.

Finally, the method was extended to select the autoregressive order when faced with ARMA disturbances. This was found to perform satisfactorily, both numerically as well as in terms of order selection.

Matlab programs to compute the CACF test and some of the examples in the paper are available from the second author.

Acknowledgments: The authors are grateful to the four editors Federico Bandi, Alex Maynard, Hyungsik Roger Moon, and Benoit Perron of this special issue in honor of Peter Phillips, and two anonymous referees, for excellent suggestions that have improved our final version of this manuscript.

Author Contributions: Both authors contributed equally to the paper.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A. Full Tables

Table A1. Simulation results of the CACF method for model *with known mean*. (The case with regressor matrix $\mathbf{X} = \mathbf{1}$ is shown in Table A2 below.) The sample size is $T = 60$. Shown here are the percentages of the 500 replications that chose AR lag length 0, 1, 2, 3, or 4, for six ARMA(2,1) models with AR parameters a_1 and a_2 and MA parameter θ . The first column, “Model”, specifies which of the six AR(2) parameter constellations and, after the decimal point, the value of the MA coefficient (e.g., 1.5 refers to AR model number 1, and the value of the MA coefficient is 0.5; similar for 1.9). Values ζ_1 and ζ_2 are the magnitude of the roots of the AR polynomial; ζ_2 missing indicates complex conjugate roots, in which case $\zeta_1 = \zeta_2$. Columns labeled “Using $\hat{\theta} = \theta$ ” assume that the MA polynomial is known; $\theta = 0$ thus corresponds to AR lag selection with no MA structure. Columns labeled “Iterative Scheme” correspond to AR lag selection when θ is not known; see Section 6 for details. **Bold faced** numbers indicate the percentage of times the true AR lag order was chosen.

AR(2) Structure					θ	Using $\hat{\theta} = \theta$					Iterative Scheme										
						AR Lag Order (%)					AR Lag Order (%)					Required Iterations (%)					
a_1	a_2	ζ_1	ζ_2			0	1	2	3	4	0	1	2	3	4	1	2	3	4	5	6+
1.0	0	0	0	.	0	81.5	4.5	4.3	5.2	4.5	86.6	0.4	5.3	3.4	4.2	85.4	11.6	1.2	0.4	0.0	1.4
2.0	1.2	−0.8	0.8944	.	0	0.0	0.0	91.4	5.0	3.6	0.6	0.0	87.0	6.1	6.4	0.5	76.7	18.7	0.5	0.0	3.5
3.0	0.7	−0.3	0.5477	.	0	0.1	35.2	56.0	4.2	4.5	71.2	6.8	15.5	3.6	3.0	68.0	19.4	7.7	0.2	0.2	4.5
4.0	0.4	−0.3	0.5477	.	0	10.7	27.0	53.5	3.7	5.1	71.4	2.7	19.6	2.9	3.3	68.5	23.0	4.0	0.2	0.2	4.0
5.0	1.4	−0.45	0.9000	0.5000	0	0.0	52.5	40.2	2.6	4.7	0.0	30.8	53.8	12.8	2.6	0.0	67.7	25.6	2.4	0.0	4.9
6.0	−0.3	0.55	−0.9066	0.6066	0	0.1	5.3	69.1	11.7	13.8	0.0	0.6	80.6	6.5	12.4	0.0	87.6	2.7	1.6	0.0	8.1
1.5	0	0	0	.	0.5	78.2	4.6	7.2	5.0	5.0	90.6	2.5	3.3	1.2	2.4	89.5	7.3	2.0	0.0	0.0	1.2
2.5	1.2	−0.8	0.8944	.	0.5	14.4	0.0	55.0	22.4	8.2	0.3	0.0	69.3	22.6	7.8	0.3	81.0	15.7	0.0	0.0	3.0
3.5	0.7	−0.3	0.5477	.	0.5	1.4	5.6	63.4	19.2	10.4	17.7	8.5	52.0	14.2	7.5	17.5	59.9	20.8	1.0	0.0	0.8
4.5	0.4	−0.3	0.5477	.	0.5	12.6	3.6	59.6	14.6	9.6	50.7	1.2	36.8	5.9	5.5	50.3	39.7	8.8	0.4	0.0	0.8
5.5	1.4	−0.45	0.9000	0.5000	0.5	0.0	15.1	68.9	9.0	7.0	0.0	20.8	73.3	4.1	1.7	0.0	82.8	15.6	0.0	0.0	1.6
6.5	−0.3	0.55	−0.9066	0.6066	0.5	0.0	0.0	24.7	54.7	20.6	3.3	0.0	33.9	41.3	21.4	2.9	70.8	11.8	1.8	0.0	12.7
1.9	0	0	0	.	0.9	76.1	5.6	7.2	5.6	5.4	80.3	4.3	7.1	3.9	4.5	79.5	18.7	0.8	0.0	0.0	1.0
2.9	1.2	−0.8	0.8944	.	0.9	0.23	0.0	61.5	24.9	13.3	0.3	0.0	58.8	27.0	14.0	0.3	90.9	7.1	0.0	0.0	1.7
3.9	0.7	−0.3	0.5477	.	0.9	4.8	8.2	51.8	19.9	15.3	5.6	6.8	45.4	23.7	18.5	5.5	83.4	8.7	0.0	0.0	2.4
4.9	0.4	−0.3	0.5477	.	0.9	24.6	1.8	44.0	15.8	13.8	26.1	1.8	37.4	16.8	17.8	25.9	65.5	7.6	0.0	0.0	1.0
5.9	1.4	−0.45	0.9000	0.5000	0.9	0.0	14.1	67.7	12.0	6.3	0.0	13.9	66.2	11.9	7.9	0.0	92.9	4.6	0.7	0.0	2.0
6.9	−0.3	0.55	−0.9066	0.6066	0.9	5.6	1.2	14.1	29.7	49.3	51.1	35.1	4.8	4.5	4.5	48.3	35.0	9.6	1.5	0.0	5.6

Table A2. Same as Table A1 but for the linear model with $X = 1$. Model numbers are now followed by an “x” to indicate the use of the regressor matrix to model the unknown mean.

	AR(2) Structure				θ	Using $\hat{\theta} = \theta$					Iterative Scheme										
	a_1	a_2	ξ_1	ξ_2		AR Lag Order (%)					AR Lag Order (%)					Required Iterations (%)					
						0	1	2	3	4	0	1	2	3	4	1	2	3	4	5	6+
1.0x	0	0	0	·	0	80.3	3.9	5.0	6.0	4.8	87.2	0.8	4.3	2.9	4.9	86.0	10.8	1.8	0.0	0.0	1.4
2.0x	1.2	−0.8	0.8944	·	0	0.0	0.0	91.0	5.3	3.7	0.5	0.0	87.5	5.7	6.3	0.5	81.4	14.6	1.3	0.0	2.1
3.0x	0.7	−0.3	0.5477	·	0	0.0	35.9	54.4	4.8	4.9	69.1	10.7	14.3	1.9	4.0	66.5	20.2	9.1	0.4	0.0	3.8
4.0x	0.4	−0.3	0.5477	·	0	10.0	28.3	53.1	4.8	3.8	76.2	2.7	16.3	1.7	3.1	73.2	19.1	3.6	0.2	0.0	3.8
5.0x	1.4	−0.45	0.9000	0.5000	0	0.0	59.9	33.0	2.9	4.2	0.0	38.2	52.9	8.8	0.0	0.0	60.5	26.3	2.6	0.0	10.5
6.0x	−0.3	0.55	−0.9066	0.6066	0	0.3	6.5	66.7	12.5	14.0	0.0	0.0	79.0	7.6	13.4	0.0	87.9	1.7	1.2	0.0	9.3
1.5x	0	0	0	·	0.5	78.8	4.8	7.0	4.6	4.8	90.8	2.0	2.9	2.0	2.3	90.0	6.9	2.2	0.0	0.0	1.0
2.5x	1.2	−0.8	0.8944	·	0.5	0.0	0.0	62.3	28.4	9.3	0.3	0.0	64.2	27.8	7.7	0.3	80.4	16.6	0.0	0.0	2.7
3.5x	0.7	−0.3	0.5477	·	0.5	1.0	7.4	61.2	22.3	8.0	15.7	12.2	48.7	17.0	6.4	15.6	59.4	23.4	1.0	0.0	0.8
4.5x	0.4	−0.3	0.5477	·	0.5	12.4	4.6	56.8	18.8	7.4	54.4	1.8	31.7	7.7	4.4	54.2	36.3	8.6	0.0	0.0	0.4
5.5x	1.4	−0.45	0.9000	0.5000	0.5	0.0	23.7	49.5	9.7	17.2	0.0	33.3	61.1	1.9	3.7	0.0	87.3	9.1	1.8	0.0	1.8
6.5x	−0.3	0.55	−0.9066	0.6066	0.5	0.0	0.0	25.8	50.3	23.9	2.9	0.0	36.2	34.6	26.2	2.5	69.9	14.0	1.2	0.0	14.4
1.9x	0	0	0	·	0.9	79.8	4.0	7.2	4.0	5.0	83.0	3.4	6.9	3.0	3.6	82.7	15.5	1.4	0.0	0.0	0.4
2.9x	1.2	−0.8	0.8944	·	0.9	0.0	0.2	59.6	27.9	12.2	0.0	0.0	59.2	28.7	12.0	0.0	91.4	5.8	0.0	0.0	2.8
3.9x	0.7	−0.3	0.5477	·	0.9	4.3	10.1	46.9	26.0	12.8	5.0	9.4	42.4	27.6	15.7	4.9	83.4	9.6	0.0	0.0	2.0
4.9x	0.4	−0.3	0.5477	·	0.9	28.1	2.6	39.1	19.4	10.8	28.7	2.6	33.0	21.1	14.6	28.2	63.6	6.6	0.0	0.0	1.6
5.9x	1.4	−0.45	0.9000	0.5000	0.9	0.0	19.4	57.3	13.6	9.7	0.0	23.9	59.7	11.9	4.5	0.0	87.0	10.1	0.0	0.0	2.9
6.9x	−0.3	0.55	−0.9066	0.6066	0.9	3.9	1.9	13.5	24.7	56.0	48.3	35.7	6.0	3.6	6.3	46.1	38.9	9.2	0.9	0.2	4.6

Table A3. Simulation results of the AIC method for model with known mean (left panel) and unknown but constant mean (right panel). Percentage of the 500 replications that chose AR lag length 0, 1, 2, 3, or 4, based on the AIC model selection criteria, for the same six ARMA(2,1) models with parameters shown in Tables A1 and A2. Columns “For AR(p) Models” ignore the MA structure present in models 1.5 to 6.5 and 1.9 to 6.9 and selects among AR(0) through AR(4); columns “For ARMA(p ,1) Models” enforces the MA structure, and compares ARMA(0,1) through ARMA(4,1); and columns “Among both Sets” uses all 10 models for comparison.

Model	No X Matrix															X = 1														
	For AR(p) Models					For ARMA(p ,1) Models					Among Both Sets					For AR(p) Models					For ARMA(p ,1) Models					Among Both Sets				
	0	1	2	3	4	0	1	2	3	4	0	1	2	3	4	0	1	2	3	4	0	1	2	3	4	0	1	2	3	4
1.0	52.2	33.8	5.2	5.4	3.4	65.8	12.0	10.0	7.2	5.0	66.0	18.6	5.0	5.6	4.8	56.2	29.4	6.2	5.2	3.0	51.0	16.8	16.0	9.2	7.0	63.4	18.6	7.4	6.0	4.6
2.0	0.0	0.0	77.0	12.8	10.2	0.2	0.0	71.0	14.8	14.0	0.2	0.0	72.0	13.4	14.4	0.0	0.0	76.2	12.6	11.2	0.2	0.0	57.2	24.4	18.2	0.0	0.0	62.4	22.2	15.4
3.0	1.2	13.2	66.6	12.8	6.2	34.4	19.8	27.6	9.4	8.8	25.2	12.0	46.0	9.2	7.6	2.2	12.0	66.8	12.0	7.0	28.4	12.6	30.6	16.6	11.8	22.4	7.6	43.6	15.4	11.0
4.0	10.8	8.6	62.6	11.0	7.0	44.6	7.0	31.0	10.8	6.6	36.4	8.6	39.4	9.0	6.6	12.8	6.6	61.4	11.0	8.2	32.6	7.2	33.6	16.8	9.8	30.2	6.8	39.0	15.4	8.6
5.0	0.0	4.4	69.8	17.0	8.8	0.0	37.2	40.8	10.6	11.4	0.0	22.8	54.2	14.4	8.6	0.0	2.6	71.8	15.8	9.8	0.0	28.0	37.0	20.8	14.2	0.0	15.6	50.0	23.0	11.4
6.0	0.0	0.2	73.8	15.8	10.2	0.0	13.8	55.0	17.2	14.0	0.0	5.8	65.6	15.6	13.0	0.0	0.6	70.4	17.8	11.2	0.0	18.2	40.8	21.8	19.2	0.0	9.4	51.2	21.0	18.4
1.5	9.6	26.0	38.2	17.0	9.2	62.6	15.8	10.4	4.6	6.6	56.4	15.4	16.0	6.2	6.0	11.4	21.6	41.2	14.2	11.6	54.2	11.4	16.4	7.0	11.0	50.0	10.2	21.6	6.8	11.4
2.5	0.0	0.0	4.4	44.8	50.8	0.4	0.6	67.8	16.2	15.0	0.4	0.6	56.8	22.8	19.4	0.0	0.0	5.2	39.0	55.8	0.4	0.2	62.8	14.8	21.8	0.4	0.2	52.2	18.8	28.4
3.5	0.0	0.0	24.0	46.8	29.2	2.6	28.2	43.6	13.6	12.0	2.4	25.0	36.4	21.6	14.6	0.0	0.0	25.2	39.2	35.6	4.4	21.0	42.4	11.4	20.8	4.2	19.4	34.4	17.4	24.6
4.5	0.2	0.6	26.6	41.8	30.8	31.2	9.0	38.8	10.0	11.0	28.8	5.8	32.8	19.6	13.0	0.2	0.6	29.8	35.0	34.4	26.0	6.2	38.2	10.0	19.6	23.6	4.4	33.4	16.0	22.6
5.5	0.0	0.2	11.2	46.4	42.2	0.0	10.0	59.0	16.8	14.2	0.0	9.0	51.0	21.8	18.2	0.0	0.2	13.2	37.6	49.0	0.0	7.2	53.0	13.4	26.4	0.0	7.0	45.0	16.6	31.4
6.5	4.2	1.0	26.8	40.8	27.2	0.4	2.4	65.8	18.2	13.2	4.0	1.8	52.6	25.8	15.8	6.0	1.6	23.4	46.2	22.8	0.4	2.2	59.2	23.6	14.6	5.8	1.6	47.0	30.6	15.0
1.9	0.2	1.2	14.2	30.4	54.0	59.4	19.4	11.2	5.4	4.6	59.4	18.6	11.2	5.8	5.0	0.4	1.0	15.8	24.4	58.4	61.4	16.4	10.0	5.8	6.4	61.4	15.8	10.0	5.8	7.0
2.9	0.0	0.0	0.0	9.8	90.2	0.0	0.6	70.2	18.0	11.2	0.0	0.6	69.8	18.2	11.4	0.0	0.0	0.2	7.2	92.6	0.0	0.4	69.0	16.8	13.8	0.0	0.4	68.6	16.8	14.2
3.9	0.0	0.0	1.4	17.6	81.0	3.8	23.8	50.2	15.4	6.8	3.8	23.8	50.0	15.2	7.2	0.0	0.0	2.0	13.2	84.8	5.2	22.6	49.6	13.8	8.8	5.2	22.6	49.6	13.8	8.8
4.9	0.0	0.0	1.8	20.6	77.6	15.4	15.0	49.0	14.2	6.4	15.4	14.8	48.8	14.4	6.6	0.0	0.0	2.0	16.4	81.6	15.4	13.0	50.2	11.8	9.6	15.4	12.8	50.2	12.0	9.6
5.9	0.0	0.0	0.2	16.4	83.4	0.0	7.2	65.2	18.4	9.2	0.0	7.0	64.6	19.2	9.2	0.0	0.6	0.6	12.0	86.8	0.0	9.8	61.4	17.2	11.6	0.0	9.8	61.2	17.4	11.6
6.9	2.2	71.2	12.4	9.2	5.0	11.4	54.2	16.4	9.2	8.8	7.2	60.6	15.6	10.2	6.4	4.0	67.2	14.2	8.4	6.2	17.2	40.8	20.2	13.4	8.4	11.2	47.0	22.4	11.8	7.6

Table A4. Simulation results of the SBC method for model with known (left panel) and unknown but constant mean (right panel). See Table A3 for comments.

Model	No X Matrix															X = 1														
	For AR(<i>p</i>) Models					For ARMA(<i>p</i> ,1) Models					Among Both Sets					For AR(<i>p</i>) Models					For ARMA(<i>p</i> ,1) Models					Among Both Sets				
	0	1	2	3	4	0	1	2	3	4	0	1	2	3	4	0	1	2	3	4	0	1	2	3	4	0	1	2	3	4
1.0	62.6	34.0	2.8	0.6	0.0	91.6	5.4	1.8	1.0	0.2	80.4	16.4	2.4	0.8	0.0	66.2	31.2	2.4	0.2	0.0	81.4	11.4	4.4	1.8	1.0	78.8	16.4	3.6	0.8	0.4
2.0	0.0	0.0	93.2	4.6	2.2	0.8	0.8	88.6	6.4	3.4	0.6	0.0	91.6	5.2	2.6	0.0	0.0	92.0	5.6	2.4	0.6	0.4	78.6	16.2	4.2	0.4	0.2	87.2	9.8	2.4
3.0	2.0	34.2	58.8	3.8	1.2	66.4	15.2	16.4	1.8	0.2	47.8	15.8	34.2	1.8	0.4	3.4	28.4	63.2	3.6	1.4	63.0	10.8	20.0	4.8	1.4	44.8	12.2	38.0	3.4	1.6
4.0	15.4	22.6	57.8	2.6	1.6	79.0	4.0	15.2	1.4	0.4	64.8	5.4	27.8	1.4	0.6	17.8	18.6	60.0	2.0	1.6	68.0	4.2	22.2	4.6	1.0	58.2	4.6	33.0	3.6	0.6
5.0	0.0	7.8	83.4	6.8	2.0	0.0	63.6	32.8	2.2	1.4	0.0	30.0	63.6	5.0	1.4	0.0	6.4	85.0	6.4	2.2	0.0	55.6	32.8	9.8	1.8	0.0	24.0	65.4	8.4	2.2
6.0	0.0	1.4	90.6	6.6	1.4	0.0	28.4	61.0	7.4	3.2	0.0	8.6	83.4	6.2	1.8	0.2	4.0	88.6	6.0	1.2	0.0	37.4	43.8	13.0	5.8	0.2	15.0	73.6	8.2	3.0
1.5	11.8	51.0	31.0	4.8	1.4	91.0	8.0	0.8	0.2	0.0	79.6	14.6	4.8	0.8	0.2	14.8	46.4	33.6	3.2	2.0	86.0	6.0	5.4	1.8	0.8	76.8	10.4	9.6	2.0	1.2
2.5	0.0	0.0	16.4	59.0	24.6	0.6	0.6	89.6	6.6	2.6	0.6	0.6	76.2	17.4	5.2	0.0	0.0	20.0	50.4	29.6	0.4	0.2	87.0	6.8	5.6	0.4	0.2	75.6	15.6	8.2
3.5	0.0	0.2	50.6	37.6	11.6	10.6	48.8	35.0	3.6	2.0	9.0	40.4	36.8	11.0	2.8	0.0	0.0	55.8	30.4	13.8	13.8	40.8	37.4	3.8	4.2	12.8	31.0	42.6	8.8	4.8
4.5	0.2	1.0	57.6	33.0	8.2	64.4	7.4	24.2	2.6	1.4	57.6	5.0	27.6	7.0	2.8	0.4	0.8	62.4	26.0	10.4	62.6	5.8	25.8	2.6	3.2	55.4	3.8	31.6	5.2	4.0
5.5	0.0	0.6	32.6	50.0	16.8	0.0	29.8	63.2	4.4	2.6	0.0	24.6	54.2	18.0	3.2	0.0	0.2	39.2	40.4	20.2	0.0	23.6	63.4	4.2	8.8	0.0	21.0	57.2	12.6	9.2
6.5	6.6	3.0	49.0	33.8	7.6	4.0	7.4	78.8	8.4	1.4	6.6	4.2	68.2	18.6	2.4	9.8	3.2	42.6	39.2	5.2	5.4	8.2	71.2	12.4	2.8	9.4	5.2	58.8	24.6	2.0
1.9	0.4	10.0	35.8	28.8	25.0	84.4	11.6	3.2	0.6	0.2	84.6	11.4	3.0	0.6	0.4	0.8	8.8	38.4	23.2	28.8	85.8	10.8	2.2	0.8	0.4	86.0	10.6	1.8	1.0	0.6
2.9	0.0	0.0	1.0	27.0	72.0	0.2	1.2	85.2	9.8	3.6	0.2	1.2	84.8	10.4	3.4	0.0	0.0	1.4	20.6	78.0	0.0	1.0	87.6	8.4	3.0	0.0	1.0	87.0	8.8	3.2
3.9	0.0	0.0	9.2	35.6	55.2	5.2	43.2	45.6	5.2	0.8	5.2	43.2	45.2	5.2	1.2	0.0	0.0	9.8	28.6	61.6	5.8	41.4	48.4	3.4	1.0	5.8	41.4	48.2	3.4	1.2
4.9	0.0	0.0	10.6	38.8	50.6	31.8	22.4	41.4	3.0	1.4	31.4	22.2	41.0	3.6	1.8	0.0	0.0	12.4	31.0	56.6	32.8	18.0	44.8	3.0	1.4	32.4	17.8	44.4	3.6	1.8
5.9	0.0	0.0	2.4	32.2	65.4	0.0	11.8	79.0	8.0	1.2	0.0	11.6	77.6	9.0	1.8	0.0	0.8	2.8	24.2	72.2	0.0	15.4	74.8	8.2	1.6	0.0	15.4	74.0	8.4	2.2
6.9	2.8	89.8	5.4	1.4	0.6	30.4	60.0	6.6	1.6	1.4	10.6	81.6	5.6	1.2	1.0	6.2	85.6	5.8	1.6	0.8	37.2	43.4	12.0	5.0	2.4	17.6	69.0	8.8	3.2	1.4

Table A5. Simulation results of the AICC method for model with known (left panel) and unknown but constant mean (right panel). See Table A3 for comments.

Model	No X Matrix															X = 1														
	For AR(<i>p</i>) Models					For ARMA(<i>p</i> ,1) Models					Among Both Sets					For AR(<i>p</i>) Models					For ARMA(<i>p</i> ,1) Models					Among Both Sets				
	0	1	2	3	4	0	1	2	3	4	0	1	2	3	4	0	1	2	3	4	0	1	2	3	4	0	1	2	3	4
1.0	53.8	34.4	5.4	4.8	1.6	71.6	12.0	8.4	5.8	2.2	69.0	18.8	5.0	5.4	1.8	58.6	30.8	5.6	3.8	1.2	58.2	16.2	14.2	7.2	4.2	66.8	18.6	7.0	5.4	2.2
2.0	0.0	0.0	80.6	11.2	8.2	0.4	0.0	76.0	13.2	10.4	0.2	0.0	77.4	12.6	9.8	0.0	0.0	81.8	10.0	8.2	0.6	0.0	64.0	23.4	12.0	0.4	0.0	68.6	19.4	11.6
3.0	1.2	15.8	67.6	10.6	4.8	40.6	20.0	26.8	7.4	5.2	28.8	13.4	46.6	6.2	5.0	2.2	14.2	70.0	9.2	4.4	35.4	13.8	29.8	12.6	8.4	25.6	8.8	46.8	11.4	7.4
4.0	11.0	9.4	64.6	10.2	4.8	52.0	6.2	29.0	7.8	5.0	40.4	7.6	39.2	7.6	5.2	13.2	7.8	65.2	8.6	5.2	42.2	6.2	34.0	11.8	5.8	35.2	6.0	40.4	12.6	5.8
5.0	0.0	4.6	74.8	13.8	6.8	0.0	41.8	40.8	9.2	8.2	0.0	24.8	57.8	10.6	6.8	0.0	3.4	77.0	13.2	6.4	0.0	33.2	39.4	18.2	9.2	0.0	17.8	55.4	19.2	7.6
6.0	0.0	0.2	78.2	13.0	8.6	0.0	16.4	58.8	14.8	10.0	0.0	6.0	70.4	12.8	10.8	0.0	0.8	76.0	14.4	8.8	0.0	21.2	43.0	21.0	14.8	0.0	10.6	57.4	17.8	14.2
1.5	9.8	31.4	39.2	13.4	6.2	69.2	15.8	9.0	2.6	3.4	60.8	16.0	14.2	5.2	3.8	11.8	27.0	43.4	10.6	7.2	61.2	12.0	14.2	5.6	7.0	57.0	11.0	18.6	6.0	7.4
2.5	0.0	0.0	6.0	48.4	45.6	0.4	0.6	74.8	14.4	9.8	0.4	0.6	62.4	22.0	14.6	0.0	0.0	6.6	42.6	50.8	0.4	0.2	71.6	13.0	14.8	0.4	0.2	58.0	19.2	22.2
3.5	0.0	0.0	28.0	48.0	24.0	3.4	33.0	45.0	10.6	8.0	3.2	28.4	37.8	19.2	11.4	0.0	0.0	32.8	41.0	26.2	5.6	26.8	44.4	9.0	14.2	4.8	23.6	37.4	16.6	17.6
4.5	0.2	0.6	31.8	42.0	25.4	37.2	10.0	37.4	7.2	8.2	32.6	6.2	34.0	17.8	9.4	0.2	0.6	36.2	35.6	27.4	32.8	7.4	37.8	7.4	14.6	29.6	4.8	35.8	13.2	16.6
5.5	0.0	0.2	13.8	49.2	36.8	0.0	13.0	61.2	15.2	10.6	0.0	11.8	52.0	21.6	14.6	0.0	0.2	18.2	40.8	40.8	0.0	9.6	60.0	10.8	19.6	0.0	9.2	50.8	16.2	23.8
6.5	4.2	1.4	31.0	42.2	21.2	0.6	4.2	70.2	16.2	8.8	4.0	2.2	57.2	25.2	11.4	6.6	2.2	27.6	46.6	17.0	0.8	3.4	65.2	21.6	9.0	6.4	2.8	51.0	29.2	10.6
1.9	0.2	1.8	16.8	33.2	48.0	65.0	19.0	9.4	4.0	2.6	65.0	18.0	9.4	4.6	3.0	0.4	1.2	19.4	27.2	51.8	69.0	16.6	8.4	2.8	3.2	68.8	16.4	8.2	2.8	3.8
2.9	0.0	0.0	0.0	12.6	87.4	0.0	0.6	73.0	17.8	8.6	0.0	0.6	72.8	18.0	8.6	0.0	0.0	0.2	10.6	89.2	0.0	0.6	75.0	15.8	8.6	0.0	0.6	74.8	15.8	8.8
3.9	0.0	0.0	2.0	21.0	77.0	4.0	27.8	52.2	12.4	3.6	4.0	27.8	52.2	12.2	3.8	0.0	0.0	2.6	16.6	80.8	5.2	26.2	52.6	11.2	4.8	5.2	26.2	52.6	11.2	4.8
4.9	0.0	0.0	2.6	24.0	73.4	15.6	16.0	51.6	12.0	4.8	15.6	16.0	51.2	12.0	5.2	0.0	0.0	2.8	19.8	77.4	17.4	15.2	52.0	10.2	5.2	17.4	15.2	52.0	10.2	5.2
5.9	0.0	0.0	0.8	20.0	79.2	0.0	7.4	69.2	16.4	7.0	0.0	7.2	68.4	17.2	7.2	0.0	0.6	0.6	14.8	84.0	0.0	10.8	66.8	15.4	7.0	0.0	10.8	66.2	15.6	7.4
6.9	2.2	74.0	11.8	8.0	4.0	13.6	57.0	16.0	7.4	6.0	7.6	65.8	14.8	7.8	4.0	4.4	73.0	11.2	7.4	4.0	21.0	43.6	18.2	11.2	6.0	12.8	52.8	18.4	10.6	5.4

Table A6. Simulation results of the HQ method for model with known (left panel) and unknown but constant mean (right panel). See Table A3 for comments.

Model	No X Matrix															X = 1														
	For AR(<i>p</i>) Models					For ARMA(<i>p</i> ,1) Models					Among Both Sets					For AR(<i>p</i>) Models					For ARMA(<i>p</i> ,1) Models					Among Both Sets				
	0	1	2	3	4	0	1	2	3	4	0	1	2	3	4	0	1	2	3	4	0	1	2	3	4	0	1	2	3	4
1.0	58.0	34.6	4.0	2.8	0.6	80.4	9.0	5.6	3.4	1.6	74.0	18.2	4.0	3.2	0.6	61.8	30.6	4.4	2.8	0.4	65.6	15.0	10.4	5.6	3.4	71.0	18.8	5.2	3.8	1.2
2.0	0.0	0.0	84.4	9.6	6.0	0.6	0.4	80.0	11.2	7.8	0.2	0.0	83.2	10.2	6.4	0.0	0.0	84.6	8.8	6.6	0.6	0.2	68.2	21.4	9.6	0.4	0.0	73.2	17.0	9.4
3.0	1.4	20.2	68.0	7.8	2.6	50.8	18.2	23.2	5.0	2.8	36.4	14.2	44.0	3.8	1.6	2.6	17.0	70.2	7.0	3.2	45.4	12.0	27.4	10.2	5.0	33.2	8.8	45.8	9.0	3.2
4.0	12.6	13.2	64.8	6.0	3.4	62.0	5.2	24.2	5.2	3.4	48.2	6.4	38.2	4.2	3.0	15.0	10.0	65.0	5.8	4.2	51.6	5.0	30.0	9.2	4.2	43.0	4.8	40.0	8.4	3.8
5.0	0.0	5.6	79.0	10.4	5.0	0.0	49.0	38.4	6.6	6.0	0.0	26.4	60.0	8.4	5.2	0.0	4.2	80.6	10.4	4.8	0.0	38.2	38.4	16.2	7.2	0.0	19.8	58.4	16.4	5.4
6.0	0.0	0.4	82.2	11.0	6.4	0.0	19.6	60.2	12.6	7.6	0.0	6.8	75.0	11.2	7.0	0.0	1.4	80.4	11.8	6.4	0.0	25.0	44.2	19.0	11.8	0.0	11.4	63.4	15.0	10.2
1.5	10.0	40.4	35.6	9.8	4.2	78.6	12.6	6.4	1.0	1.4	70.0	16.0	9.2	2.8	2.0	13.0	35.2	39.6	7.2	5.0	70.6	10.2	10.8	4.0	4.4	64.4	12.0	15.8	3.2	4.6
2.5	0.0	0.0	7.8	52.0	40.2	0.6	0.6	79.0	12.0	7.8	0.6	0.6	67.2	20.0	11.6	0.0	0.0	9.4	45.6	45.0	0.4	0.2	77.4	10.6	11.4	0.4	0.2	64.6	17.4	17.4
3.5	0.0	0.0	34.4	45.4	20.2	4.2	39.0	41.8	9.2	5.8	3.8	32.8	38.8	16.6	8.0	0.0	0.0	38.2	39.2	22.6	7.0	30.6	43.2	7.8	11.4	6.6	25.0	40.2	14.8	13.4
4.5	0.2	0.6	41.2	39.8	18.2	46.6	9.4	33.6	5.0	5.4	40.8	6.4	33.8	13.4	5.6	0.2	0.6	44.6	32.8	21.8	42.6	7.0	33.8	5.6	11.0	37.6	5.0	34.4	11.2	11.8
5.5	0.0	0.4	18.8	50.2	30.6	0.0	18.4	62.4	11.2	8.0	0.0	16.4	53.0	21.0	9.6	0.0	0.2	22.6	41.8	35.4	0.0	14.0	59.4	9.0	17.6	0.0	12.8	50.8	15.6	20.8
6.5	5.2	2.2	36.2	40.0	16.4	1.2	5.0	75.0	13.0	5.8	5.0	3.4	60.8	23.2	7.6	7.2	2.4	32.2	44.2	14.0	2.0	5.0	67.0	18.2	7.8	7.2	3.4	53.8	27.2	8.4
1.9	0.2	3.8	20.0	33.6	42.4	73.8	16.0	6.8	1.8	1.6	73.8	16.0	6.4	2.0	1.8	0.6	3.0	22.8	27.6	46.0	74.8	14.4	7.0	1.8	2.0	74.4	14.2	7.2	1.8	2.4
2.9	0.0	0.0	0.2	16.8	83.0	0.0	0.8	77.6	15.2	6.4	0.0	0.8	77.2	15.6	6.4	0.0	0.0	0.2	13.2	86.6	0.0	0.6	78.2	13.6	7.6	0.0	0.6	78.0	13.6	7.8
3.9	0.0	0.0	3.8	25.0	71.2	4.6	32.2	50.2	10.2	2.8	4.6	32.2	50.2	10.0	3.0	0.0	0.0	3.8	19.4	76.8	5.4	30.0	52.2	8.2	4.2	5.4	30.0	52.2	8.2	4.2
4.9	0.0	0.0	4.2	28.0	67.8	19.4	19.0	50.0	8.0	3.6	19.4	18.8	49.6	8.2	4.0	0.0	0.0	4.8	22.6	72.6	20.0	16.2	51.6	8.2	4.0	20.0	16.2	51.2	8.4	4.2
5.9	0.0	0.0	0.8	22.6	76.6	0.0	8.6	72.0	14.2	5.2	0.0	8.2	71.4	14.8	5.6	0.0	0.8	0.8	17.4	81.0	0.0	11.4	68.0	15.2	5.4	0.0	11.4	67.0	15.4	6.2
6.9	2.6	80.0	8.8	6.0	2.6	18.8	60.0	12.2	5.6	3.4	9.0	72.0	9.8	6.2	3.0	5.2	78.2	8.2	5.8	2.6	25.0	43.4	16.6	10.6	4.4	14.6	58.0	14.4	9.0	4.0

Table A7. Simulation results of the FPE method for model with known (left panel) and unknown but constant mean (right panel). See Table A3 for comments.

Model	No X Matrix															X = 1														
	For AR(<i>p</i>) Models					For ARMA(<i>p</i> ,1) Models					Among both Sets					For AR(<i>p</i>) Models					For ARMA(<i>p</i> ,1) Models					Among both Sets				
	0	1	2	3	4	0	1	2	3	4	0	1	2	3	4	0	1	2	3	4	0	1	2	3	4	0	1	2	3	4
1.0	52.2	33.8	5.2	5.4	3.4	66.2	12.2	9.8	7.2	4.6	66.0	18.8	5.0	5.6	4.6	56.2	29.4	6.2	5.2	3.0	51.0	16.8	16.0	9.2	7.0	63.4	18.6	7.4	6.0	4.6
2.0	0.0	0.0	77.0	12.8	10.2	0.2	0.0	71.0	14.8	14.0	0.2	0.0	72.0	13.4	14.4	0.0	0.0	76.2	12.6	11.2	0.2	0.0	57.4	24.4	18.0	0.0	0.0	62.4	22.2	15.4
3.0	1.2	13.2	66.6	12.8	6.2	34.4	19.8	27.8	9.4	8.6	25.2	12.0	46.2	9.2	7.4	2.2	12.0	67.0	12.2	6.6	28.6	12.6	30.6	16.4	11.8	22.4	7.6	44.0	15.2	10.8
4.0	10.8	8.6	62.6	11.0	7.0	44.6	7.0	31.0	10.8	6.6	36.4	8.6	39.4	9.0	6.6	12.8	6.6	61.4	11.0	8.2	32.8	7.2	33.6	16.6	9.8	30.2	6.8	39.0	15.4	8.6
5.0	0.0	4.4	69.8	17.0	8.8	0.0	37.2	40.8	10.6	11.4	0.0	22.8	54.2	14.4	8.6	0.0	2.6	72.0	15.8	9.6	0.0	28.2	37.2	20.8	13.8	0.0	15.8	50.4	22.8	11.0
6.0	0.0	0.2	73.8	15.8	10.2	0.0	13.8	55.2	17.0	14.0	0.0	5.8	65.8	15.6	12.8	0.0	0.6	70.4	17.8	11.2	0.0	18.2	40.8	22.0	19.0	0.0	9.4	51.2	21.0	18.4
1.5	9.6	26.0	38.2	17.0	9.2	62.6	15.8	10.4	4.6	6.6	56.4	15.4	16.0	6.2	6.0	11.4	21.6	41.4	14.2	11.4	54.4	11.4	16.6	7.0	10.6	50.0	10.2	21.8	6.8	11.2
2.5	0.0	0.0	4.4	45.0	50.6	0.4	0.6	67.8	16.2	15.0	0.4	0.6	57.0	22.8	19.2	0.0	0.0	5.2	39.0	55.8	0.4	0.2	63.2	15.0	21.2	0.4	0.2	52.4	18.8	28.2
3.5	0.0	0.0	24.2	46.8	29.0	2.6	28.4	43.6	13.8	11.6	2.4	25.2	36.4	21.6	14.4	0.0	0.0	25.2	39.6	35.2	4.4	21.0	42.6	11.4	20.6	4.2	19.4	34.6	17.6	24.2
4.5	0.2	0.6	26.8	41.8	30.6	31.2	9.0	38.8	10.0	11.0	28.8	5.8	32.8	19.6	13.0	0.2	0.6	30.0	34.8	34.4	26.2	6.2	38.2	10.0	19.4	23.6	4.4	33.4	16.0	22.6
5.5	0.0	0.2	11.2	46.6	42.0	0.0	10.0	59.4	16.6	14.0	0.0	9.0	51.2	21.6	18.2	0.0	0.2	13.4	37.8	48.6	0.0	7.2	53.2	13.2	26.4	0.0	7.0	45.4	16.4	31.2
6.5	4.2	1.0	26.8	40.8	27.2	0.4	2.4	66.0	18.0	13.2	4.0	1.8	52.6	25.8	15.8	6.0	1.6	23.4	46.2	22.8	0.4	2.2	59.4	23.6	14.4	5.8	1.6	47.2	30.6	14.8
1.9	0.2	1.2	14.2	30.6	53.8	59.4	19.4	11.2	5.4	4.6	59.4	18.6	11.2	5.8	5.0	0.4	1.0	15.8	24.6	58.2	61.4	16.4	10.0	5.8	6.4	61.4	15.8	10.0	5.8	7.0
2.9	0.0	0.0	0.0	9.8	90.2	0.0	0.6	70.6	17.8	11.0	0.0	0.6	70.2	18.0	11.2	0.0	0.0	0.2	7.2	92.6	0.0	0.4	69.2	16.8	13.6	0.0	0.4	68.8	16.8	14.0
3.9	0.0	0.0	1.4	17.6	81.0	3.8	23.8	50.2	15.4	6.8	3.8	23.8	50.0	15.2	7.2	0.0	0.0	2.0	13.2	84.8	5.2	22.6	49.8	13.6	8.8	5.2	22.6	49.8	13.6	8.8
4.9	0.0	0.0	1.8	20.8	77.4	15.4	15.0	49.0	14.2	6.4	15.4	14.8	48.8	14.4	6.6	0.0	0.0	2.0	16.4	81.6	15.4	13.2	50.4	11.6	9.4	15.4	13.0	50.4	11.8	9.4
5.9	0.0	0.0	0.2	16.4	83.4	0.0	7.2	65.4	18.4	9.0	0.0	7.0	64.8	19.2	9.0	0.0	0.6	0.6	12.2	86.6	0.0	9.8	61.4	17.4	11.4	0.0	9.8	61.2	17.6	11.4
6.9	2.2	71.2	12.4	9.2	5.0	11.4	54.6	16.2	9.2	8.6	7.2	60.6	15.6	10.2	6.4	4.0	67.4	14.2	8.4	6.0	17.2	40.8	20.2	13.4	8.4	11.2	47.2	22.4	11.6	7.6

References

- Abadir, Karim M., and Jan R. Magnus. 2005. *Matrix Algebra*. Cambridge: Cambridge University Press.
- Anderson, Theodore W. 1971. *The Statistical Analysis of Time Series*. New York: John Wiley & Sons.
- Andrews, Donald W. K. 1993. Exactly Median-Unbiased Estimation of First Order Autoregressive/Unit Root Models. *Econometrica* 61: 139–65.
- Andrews, Donald W. K., and Patrik Guggenberger. 2014. A Conditional-Heteroskedasticity-Robust Confidence Interval for the Autoregressive Parameter. *The Review of Economics and Statistics* 96: 376–81.
- Barndorff-Nielsen, O. E., and D. R. Cox. 1979. Edgeworth and Saddlepoint Approximations with Statistical Applications (with discussion). *Journal of the Royal Statistical Society, Series B* 41: 279–312.
- Brockwell, Peter J., and Richard A. Davis. 1991. *Time Series: Theory and Methods*, 2nd ed. Berlin: Springer.
- Broda, Simon A., Kai Carstensen, and Marc S. Paoletta. 2007. Bias—Adjusted Estimation in the ARX(1) Model. *Computational Statistics & Data Analysis* 51: 3355–67.
- Broda, Simon A., Kai Carstensen, and Marc S. Paoletta. 2009. Assessing and Improving the Performance of Nearly Efficient Unit Root Tests in Small Samples. *Econometric Reviews* 28: 468–94.
- Brooks, Stephen P., Paolo Giudici, and G. O. Roberts. 2003. Efficient Construction of Reversible Jump Markov Chain Monte Carlo Proposal Distributions (with discussion). *Journal of the Royal Statistical Society, Series B* 65: 3–55.
- Burnham, Kenneth P., and David R. Anderson. 2003. *Model Selection and Multi-Model Inference*, 2nd ed. New York: Springer.
- Butler, Ronald W. 2007. *An Introduction to Saddlepoint Methods*. Cambridge: Cambridge University Press.
- Butler, Ronald W., and Marc S. Paoletta. 1998. Approximate Distributions for the Various Serial Correlograms. *Bernoulli* 4: 497–518.
- Chatfield, Christopher. 2001. *Time-Series Forecasting*. Boca Raton: Chapman & Hall.
- Chib, Siddhartha. 1993. Bayes Estimation of Regressions with Autoregressive Errors: A Gibbs Sampling Approach. *Journal of Econometrics* 58: 275–94.
- Choi, Byoung-Seon. 1992. *ARMA Model Identification*. New York: Springer.
- Daniels, H. E. 1954. Saddlepoint Approximation in Statistics. *The Annals of Mathematical Statistics* 25: 631–50.
- Daniels, H. E. 1956. The Approximate Distribution of Serial Correlation Coefficients. *Biometrika* 43: 169–85.
- Daniels, H. E. 1987. Tail Probability Approximation. *International Statistical Review* 55: 37–48.
- Daniels, H. E., and G. A. Young. 1991. Saddlepoint Approximation for the Studentized Mean, with an Application to the Bootstrap. *Biometrika* 78: 169–79.
- Dubbelman, C., A. S. Louter, and A. P. J. Abrahamse. 1978. On Typical Characteristics of Economic Time Series and the Relative Qualities of Five Autocorrelation Tests. *Journal of Econometrics* 8: 295–306.
- Durbin, J. 1980. The Approximate Distribution of Serial Correlation Coefficients Calculated from Residuals on Fourier Series. *Biometrika* 67: 335–50.
- Durbin, J., and G. S. Watson. 1950. Testing for Serial Correlation in Least Squares Regression. I. *Biometrika* 37: 409–28.
- Durbin, J., and G. S. Watson. 1971. Testing for Serial Correlation in Least Squares Regression. III. *Biometrika* 58: 1–19.
- Elliott, Graham, and James H. Stock. 2001. Confidence Interval for Autoregressive Coefficients Near One. *Journal of Econometrics* 103: 155–81.
- Field, Christopher, and Elvezio Ronchetti. 1990. *Small Sample Asymptotics. Lecture Notes—Monograph Series, Volume 13*. Hayward: Institute of Mathematical Statistics.
- Fomby, Thomas B., and David K. Guilkey. 1978. On Choosing the Optimal Level of Significance for the Durbin–Watson Test and the Bayesian Alternative. *Journal of Econometrics* 8: 203–13.
- Granger, Clive W. J., Maxwell L. King, and Halbert White. 1995. Comments on Testing Economic Theories and the Use of Model Selection Criteria. *Journal of Econometrics* 67: 173–87.
- Han, Chirok, Peter C. B. Phillips, and Donggyu Sul. 2017. Lag Length Selection in Panel Autoregression. *Econometric Reviews* 36: 225–40.
- Holly, Alberto, and Peter C. B. Phillips. 1979. A Saddlepoint Approximation to the Distribution of the k -Class Estimator of a Coefficient in a Simultaneous System. *Econometrica* 47: 1527–47.
- Jenkins, Gwilym M., and Athar S. Alevi. 1981. Some Aspects of Modelling and Forecasting Multivariate Time Series. *Journal of Time Series Analysis* 2: 1–47.

- Jensen, Jens L. 1995. *Saddlepoint Approximations*. Oxford: Oxford University Press.
- Keuzenkamp, Hugo A., and Michael McAleer. 1997. The Complexity of Simplicity. In *11th Biennial Conference on Modelling and Simulation*. Edited by P. Binning, H. Bridgman, M. McAleer and B. Williams. Special issue, Mathematics and Computers in Simulation 43: 553–61. Available online: <http://www.sciencedirect.com/science/article/pii/S037847549700044X> (accessed on 8 May 2017).
- King, Maxwell L. 1985. A Point Optimal Test for Autoregressive Disturbances. *Journal of Econometrics* 27: 21–37.
- King, Maxwell L., and Sivagowry Srikanthakumar. 2015. Point Optimal Testing: A Survey of the Post 1987 Literature. *Model Assisted Statistics and Applications* 10: 179–96.
- Kolassa, John E. 1996. Higher-Order Approximations to Conditional Distribution Functions. *Annals of Statistics* 24: 353–64.
- Kolassa, John E. 2006. *Series Approximation Methods in Statistics*, 3rd ed. New York: Springer.
- Konishi, Sadanori, and Genshiro Kitagawa. 2008. *Information Criteria and Statistical Modeling*. New York: Springer.
- Koreisha, Sergio, and Gary Yoshimoto. 1991. A Comparison among Identification Procedures for Autoregressive Moving Average Models. *International Statistical Review* 59: 37–57.
- Le, Nhu D., Adrian E. Raftery, and R. Douglas Martin. 1996. Robust Bayesian Model Selection for Autoregressive Processes with Additive Outliers. *Journal of the American Statistical Association* 91: 123–31.
- Leeb, Hannes, and Benedikt M. Pötscher. 2005. Model Selection and Inference: Facts and Fiction. *Econometric Theory* 21: 21–59.
- Lehmann, E. L. 1958. Significance Level and Power. *The Annals of Mathematical Statistics* 29: 1167–76.
- Lehmann, E. L. 1986. *Testing Statistical Hypotheses*, 2nd ed. New York: John Wiley & Sons.
- Lugannani, Robert, and Stephen O. Rice. 1980. Saddlepoint Approximations for the Distribution of Sums of Independent Random Variables. *Advances in Applied Probability* 12: 475–90.
- Lütkepohl, Helmut. 2005. *New Introduction to Multiple Time Series Analysis*. Berlin: Springer.
- Makridakis, Spyros, and Michele Hibon. 2000. The M3 Competition: Results, Conclusions and Implications. *International Journal of Forecasting* 17: 567–70.
- McGregor, J. R. 1960. An Approximate Test for Serial Correlation in Polynomial Regression. *Biometrika* 47: 111–19.
- McLeod, Ian. 1975. Derivation of the Theoretical Autocovariance Function of Autoregressive–Moving Average Time Series. *Applied Statistics* 24: 255–56. Correction: 1977, 26: 194.
- McQuarrie, Allan D. R., and Chih-Ling Tsai. 1998. *Regression and Time Series Model Selection*. River Edge: World Scientific.
- Mittnik, Stefan. 1988. Derivation of the Theoretical Autocovariance and Autocorrelation Function of Autoregressive Moving Average Processes. *Communications in Statistics—Theory and Methods* 17: 3825–31.
- Pankratz, Alan. 1983. *Forecasting with Univariate Box–Jenkins Models: Concepts and Cases*. New York: John Wiley & Sons.
- Phillips, Peter C. B. 1978. Edgeworth and Saddlepoint Approximations in a First Order Non-Circular Autoregression. *Biometrika* 65: 91–98.
- Phillips, Peter C. B. 2008. Unit Root Model Selection. *Journal of the Japan Statistical Society* 38: 65–74.
- Phillips, Peter C. B. 2014. On Confidence Intervals for Autoregressive Roots and Predictive Regression. *Econometrica* 82: 1177–95.
- Ploberger, Werner, and Peter C. B. Phillips. 2003. Empirical Limits for Time Series Econometric Models. *Econometrica* 71: 627–73.
- Pötscher, B. M. 1983. Order Estimation in ARMA Models by Lagrangian Multiplier Tests. *The Annals of Statistics* 11: 872–85.
- Rahman, M. S., and Maxwell L. King. 1999. Improved Model Selection Criterion. *Communications in Statistics—Simulation and Computation* 28: 51–71.
- Reid, N. 1988. Saddlepoint Methods and Statistical Inference (with discussion). *Statistical Science* 3: 213–38.
- Sanathanan, Lalitha. 1974. Critical Power Function and Decision Making. *Journal of the American Statistical Association* 69: 398–402.
- Schervish, M. J., and R. S. Tsay. 1988. Bayesian Modeling and Forecasting in Autoregressive Models. In *Bayesian Analysis of Time Series and Dynamic Models*. Edited by J. C. Spall. New York: Marcel Dekker, pp. 23–52.
- Skovgaard, Ib M. 1987. Saddlepoint Expansions for Conditional Distributions. *Journal of Applied Physiology* 24: 275–87.
- Tiao, G. C., and G. E. P. Box. 1981. Modelling Multiple Time Series with Applications. *Journal of the American Statistical Association* 76: 802–16.

- Van der Leeuw, Jan. 1994. The Covariance Matrix of ARMA Errors in Closed Form. *Journal of Econometrics* 63: 397–405.
- Zellner, Arnold. 1971. *An Introduction to Bayesian Inference in Econometrics*. New York: John Wiley & Sons.
- Zellner, Arnold. 2001. Keep it Sophisticatedly Simple. In *Simplicity, Inference and Modelling*. Edited by A. Zellner, H. A. Keuzenkamp and M. McAleer. Cambridge: Cambridge University Press, pp. 242–62.



© 2017 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).